THE UNIVERSITY
of EDINBURGH

CRISPR/Cas9-based Strategies for Pig Host Resistance to Influenza A Virus

Yue Du

Submitted for the degree of Doctor of Philosophy

The University of Edinburgh

January 2019

To my mother

&

In the love memory of my father

Acknowledgement

I definitely have a long important list of people deserving of great thanks.

Firstly, I would like to express my sincere gratitude appreciation to my principle supervisor, Prof. Bruce Whitelaw, for his continuous support of my PhD study. The guidance, patience, motivation, and encouragement helped me get through my PhD life. I would also like to extend my deepest gratitude to my second supervisor, Dr. Simon Lillico. It has been such a pleasure to learn from him, who is one of the smartest and kindest scientists. Thanks for guiding me to grow as a researcher. I could not have imaged having a better advisor for my Ph.D study. I would like to thank my committee members, Prof. Dutia Bernadette and Prof. Paul Digard. In particular, Paul is super patient and genuine to me, and willing to get me out from any problems and upset emotions. I can never completely express my appreciation to Paul.

Besides my advisors, I would like to express my sincere thanks to the examiners of the final viva, Dr. Robert Dalziel and Dr. Alun Luo, for the insightful comments and critics on the thesis. I enjoy my defence as it was such an enjoyable and important moment in my life. Thanks should also go to Prof. Denis Headon, for his encouragement, attention and assistance. ‘Hard work, and success follows’ is what he encouraged me through the toughest time, and these are what I will never forget.

I had a great pleasure of working with the past and current members in the Whitelaw Group: Dr. Chris proudfoot, Claire Neil, Dr. Vrushali Patil, Dr. Alex Brown, Dr. Maeve Ballantyne, Dr. Sarah Fletcher, Dr. Akshay Joshi, Dr. Sylvia Beka, Dr. Spring Tan, Gus Mcfarlane, and Hamish Salvesen. Many thanks for their support, advice, and all the fun we have had in the last three years. Thanks for making the Whitelaw group a sweet family. I was also grateful and honoured to be involved in I&I group. Though countless times of climbing

from first floor to the second floor throughout the time there for the virology work, I was definitely enjoying copious amounts of hilarity and scientific atmosphere. Thanks, Rute Maria Pinto, Dr. Seema Jasim, Dr. Liliane Chung, Dr. Christine Tait-Burkard, Carina Conceição, Susi Keane. They are enormously important to me.

I am indebted to all family in Edinburgh who opened their sweet home to me during my time in Edinburgh, Dr. Debiao Zhao and and Ms. Suling Zhao, Dr. Zhiguang Wu and Dr. Jun Wu. They were always managed to help me out in numerous way. Also, I would like to give thanks to my friends at the Roslin Institute: Dr. Alewo Akoh, Dr. Chandana Basu, Dr. Gwen Tang, Dr. William Ho, James Ozanne, Dr. Greg Markby, Dr. Omar Alfituri, Ciara A Farren. Thanks for all the blessings and accompanies. Many thanks to great friends in Edinburgh, Hongjiang Wu, Yan Zhu, Xueyi Shen, Yuanyuan Liu, Chong Chao, Jingjing Gong, Xiangrui Meng, Mosi Li, and Jianpeng Cheng, for their friendship and all the wonderful memories. A heartfelt thank you goes to Dr. Michael Bai, who has been by my side, for his enormous support and encouragement along the journey. I am so proud of having you in my life.

Last but not the least, I would like to express my special appreciation to Pan, for everything in general, especially his unwavering support and patience throughout my Ph.D life. I am extremely grateful to my beloved mother, Ms. Yu, for her strong and unconditional love; for putting up with my own particular brand of stubborn; and for the strong belief in me. Your love for me was what sustained me thus far.

Now you believe, there are too many people I would like to express my appreciation to. I had an incredible PhD life in the unprecedented learning environment. I accept that all great moments must come to the archive (as the memories, BUT NOT THE END), but realistically I think I will keep a strong faith on **SCIENCE** and **CARRY ON**.

Declaration

I hereby declare that this thesis presents my own myself unless otherwise stated. The work has not been submitted for any other degree or professional qualification.

Print name: Yue Du

Signature:

Date:

Lay Summary

Influenza A virus (IAV) is a highly infectious pathogen circulating in avian and mammalian species. Pigs are susceptible to both human and avian IAVs, owing to the existence of, SA-α 2,6-Gal and SA α 2,3-Gal-containing host cell virus receptors. This enables pigs to serve as intermediate hosts in the production of reassortant IAV strains, which facilitates cross-species transmission, and can lead to epidemics and pandemics. Due to the rapid antigenic evolution, traditional vaccines and licensed antiviral drugs are of limited effectiveness for controlling pandemic causing by newly emerging and re-emerging IAV strains. In this work, we apply CRISPR/Cas9-based approaches to engineer the pig ST6Gal1 coding gene, which impairs the biosynthesis of α 2,6-sialylated glycan structure exclusively on the airway cells, and thus reduces IAV entry. Moreover, the normal biological functions conducted by α 2,6-sialylated glycan receptors may be retained. We propose this work serves as proof of principle for generating genetic edited pigs for IAV host resistance as a strategy to achieve improvement in pig herds.

Abstract

Influenza circulates in different mammalian and avian species, causing epidemics and occasional pandemics. This poses a substantial threat to agricultural productions, animal welfare, human public health, and economy. The causative agent of the disease is Influenza A virus (IAV), whose entry depends on its preference on the receptor molecules since the preference determines whether virus glycoproteins can employ the host cell surface sialic acid (SA) as ligands. There are two major types of sialylated glycans as receptors for viral recognition: one is Neu5Ac α 2,6-Gal (SA α 2,6-Gal for short) preferentially recognised by human IAVs, and the other one is Neu5Ac α 2,3-Gal (SA α 2,3-Gal for short) predominantly recognised by avian IAVs. Pigs harbouring both types of SA-containing receptors have the potential to play a role of 'intermediate hosts' or 'mixing vessels'. Therefore, it is of great interest to develop methods aiming at reducing α 2,6-SA-containing receptors in pig cells, so that they are less susceptible to human IAV infections.

β-galactoside α 2,6-sialyltransferase 1 (ST6Gal1) mediates N-linked α 2,6-sialylation on cell surfaces by catalysing the addition of α 2,6-SA to the terminal N-glycans. ST6Gal1 is involved in a wide range of biological events, such as the generation of carbohydrate determinants on the cell surfaces, the immune regulation, and in various carcinomas. ST6Gal1 is encoded by the ST6GAL1 gene, which expression has been reported to display a tissue-specific pattern as a result of the regulations of multiple promoter regions and the differential combinations of 5'-untranslated exons.

Driven by the concern that inactivating the ST6GAL1 gene may have deleterious phenotypic effects given the widespread expression profile and the

diversity of its biological functions, I pursued a subtle approach to engineering the ST6GAL1 gene — that of removing a single promoter region to alter the expression profile. I hypothesised that reducing the biosynthesis of α 2,6-sialylated glycan structure exclusively on the respiratory tract could potentially block IAV entry without compromising humoral immune responses.

To this end, I identified 5' transcription starting sites (TSSs) and 5' untranslated region (UTR) exons of transcripts expressed in nine pig tissues. Then we employed the CRISPR/Cas9 system to precisely engineer pig ST6GAL1 gene instead of the whole gene deletion. The consequence of deleting the region surrounding the 5' TSS of ST6GAL1 transcripts predominantly expressed in airway was assessed (the resulting model was termed as ST6GAL1ΔP). Moreover, I generated a ST6GAL1 functional knockout model by inducing a frameshift mutation in pig trachea cells (the resulting model was termed as ST6GALΔCD). Additionally, human IAV had reduced infectivity in ST6GAL1ΔP relative to non-edited cells, suggesting that a strategy to reduce the biosynthesis of α 2,6-sialylated glycan structure exclusively on the airway could offer an antiviral strategy, independent of inducing a humoral immune response.

This work lays a solid foundation in generating engineered pigs for IAV host resistance modelling, and helps us to achieve the genetic improvement in swine herds; also, it provides a good understanding of the fundamental molecular basis of the IAV-host interactions, and develop novel antiviral and therapeutics strategies.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

-/-	Homozygous
-/+	Heterozygous
Δ	Deletion
5' RACE	5' Rapid Amplification of cDNA Ends
AAP	Abridged Anchor primer
AmpR	Ampicillin Resistance
ATP	Adenosine Triphosphate
AUAP	Abridged Universal Amplification Primer
BCP	1-Bromo-3-Chloropropane
bp	Base pair
BSA	Bovine Serum Albumin
Cas9	CRISPR-associated protein 9
Cat#	Catalogue
cDNA	Complementary Deoxyribonucleic acid
CMP	Cytidine 5′-monophosphate
CMP-SA	Cytidine 5′-monophosphate- Sialic acid transporter
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeat
cRNA	Complementary Ribonucleic acid
cRNP	Complementary Ribonucleoprotein
crRNA	CRISPR RNA
Ct	Threshold Cycle
DAPI	4',6-Diamidino-2-Phenylindole
dCTP	Deoxycytidine Triphosphate
ddH_2O	Deionised water
D-MEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	DNA Triphosphate
dNTP	Deoxynucleotide
DPE	Downstream Promoter Element
DSB	Double Stranded Break

DSBR	Double Stranded Break Repair
dsDNA	Double-stranded Deoxyribonucleic Acid
DTT	Dithiothreitol
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
FACs	Fluorescent-Activated Cell sorting
FBS	Fetal Bovine Serum
FSC	Forward Scatter
GADPH	Glyceraldehyde 3-Phosphate Dehydrogenase
Gal	Galactose
GFP	Green Fluorescent Protein
GlaNAc	N-acetyl-galactosamine
GlcNAc	N-acetyl-glucosamine
GM	Genetically Modified
GSP	Gene-Specific Primer
h.p.i	Hour post infection
h.p.t	Hour post transfection
HA	Haemagglutinin
HDR	Homology Directed Repair
HMW	High Molecular Weight
HPAIV	High Pathogenic Avian Influenza Virus
HR	Homologous Recombination
IAV	Influenza A virus
IBV	Influenza B virus
ICV	Influenza C virus
IF	Immunofluorescence
IIV	Inactivated Influenza Vaccine
in/dels	insertions and deletions
kDa	Kilo Dalton
KDN	2-Keto-3-Deoxynononic acid
KO	Knockout
LAIV	Live Attenuated Influenza Vaccine
LPAIV	Low Pathogenic Avian Influenza Virus

M1	Matrix Protein 1
M2	Matrix Protein 2
M.O.I	Multiplicity of Infection
MDCK	Madin-Darby Canine Kidney
MDCK-SIAT1	MDCK with the cDNA of human ST6Gal1
MMEJ	Microhomology-mediated End Joining
mRNA	Messenger Ribonucleic acid
NA	Neuraminidase
NaCl	Sodium Chloride
NBF	Neutral Buffered Formalin
Neu5Ac	N-acetyl-neuraminic acid
Neu5Gc	N-glycolyl-neuraminic acid
NF	Nuclease-free
NHEJ	Non-homologous End Joining
NLS	Nuclear localisation signal
NP	Nucleoprotein
NPC	Nuclear pore complex
NPTr	Newborn pig trachea
NS2	Non-structural protein
nt	Nucleotide
Ori	Origin of replication
PA	Polymerase acidic protein
PAM	Protospacer Adjacent Motif
PAM	Protospacer Adjacent Motif
PB 1	Polymerase basic protein 1
PB 2	Polymerase basic protein 2
PBS	Phosphate buffered saline
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
Pen/Strep	Penicillin and Streptomycin
PK15	Pig Kidney 15
Pol II	DNA-dependent RNA Polymerase II
PR8	Influenza A/Puerto Rico/8/34

QIV	Quadrivalent Influenza Vaccine
RBS	Receptor binding site
RdRp	DNA-dependent RNA polymerase II
RIN	RNA Integrity Number
RNA	Ribonucleic Acid
RNase	Ribonuclease
RNP	Ribonucleoprotein
RPL4	Ribosomal protein L4
RT	Reverse Transcriptase
SD	Standard Deviation
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEM	Standard Error of the Mean
SFM	Serum-free Medium
sgRNA	Single guide RNA
***Sp*Cas9**	*Streptococcus Pyogenes* Cas9
SSC	Side scatter
ssODN	Single-stranded donor oligonucleotides
ST	Sialyltransferase
TAE	Tris-acetate-EDTA
TALE	Transcription Activator-Like Effector
TALEN	Transcription Activator-Like Effector Nuclease
TdT	Terminal deoxynucleotidyl transferase
TdT	Terminal deoxynucleotidyl transferase
TE	TrypleTM Express Enzyme
TE	Tris-EDTA
TG	Transgenic
TIDE	Tracking in/dels by decomposition
TIV	Inactivated Trivalent Influenza Vaccine
TPCK	6- (1-tosylamido-2-phenyl) ethyl chloromethyl keton
TracrRNA	Trans-activating CRISPR RNA
Tris-HCL	Tris Hydrochloride
TSS	Transcription start site

UAP	Universal Amplification Primer
UTE	Untranslated Exon
UTR	Untranslated region
V	Voltage
v/v	Volume per volume
VGM	Viral Growth Medium
vRNA	Virion RNA
vRNP	Viral Ribonucleoprotein complex
w/v	Weight per volume
WB	Western blot/blotting
WT	Wild type
ZF	Zinc-finger
ZFN	Zinc-Finger Nucleases
ZFP	Zinc-finger protein

Ethics statement

All animal procedures are performed at the University of Edinburgh, UK. All samples are taken from animals culled ethically under the 3Rs (**R**eplacement, **R**eduction and **R**efinement).

Preface

In this thesis, I utilise CRISPR/Cas9 system-based strategies to precisely engineer host receptors against IAV entry. The thesis is organised as follows. **Chapter 1** presents a general literature review. All the materials and methods used in this project are then described in **Chapter 2**. The target gene ST6GAL1, its tissue-abundant expression patterns, and its associated promoters are discussed in **Chapter 3**. Then, in **Chapters 4**, I evaluate strategies for precisely engineering the ST6GAL1 gene in pig trachea cells using the CRISPR/Cas9 system. In **Chapter 5**, the sialylation levels on the surfaces of edited cells and their susceptibility to IAV infection are assessed. **Chapter 7** concludes this work and discusses potential future directions.

1 General introduction

1.1 General introduction to Influenza A virus

1.1.1 Taxonomy

Influenza is a contagious and acute disease circulating in mammalian and avian species. There are four influenza virus (IV) genera, influenza A, B, C and D, termed as IAV, IBV, ICV and D, respectively. These four genera are different in their morphology, surface glycoproteins, antigenic viral encoded matrix (M) and nucleoproteins (NP), and genome segments (IA/BV contains eight, whereas ICV has seven) (Dowdle, 2006). IAV circulates in a wide range of host species, such as humans, avian, horses, pigs, dogs, and marine mammals. It is associated with the emergence of annual flu epidemics, as well as the occasional pandemics in the human population (Michaelis et al., 2009); IBV is only found in humans and domestic pigs (Ran et al., 2015); ICV was reported to have little clinical importance (Matsuzaki et al., 2006). IDV contains seven negative single-stranded RNA segments. IDV has 50% identity of amino acid sequence with those in ICV. IDV was isolated from pigs and then cattle. (Su et al., 2017). Therefore, among these four genera, IAV is the most important concern in public health.

1.1.2 Virion structure

IAV belongs to the *Orthomyxoviridae* family. It is pleomorphic, enveloped, and spherical or filamentous, with the diameter 80 to 120 nm (Pleschka, 2012). The structure of one single viral particle in shown in **Figure 1.1.**

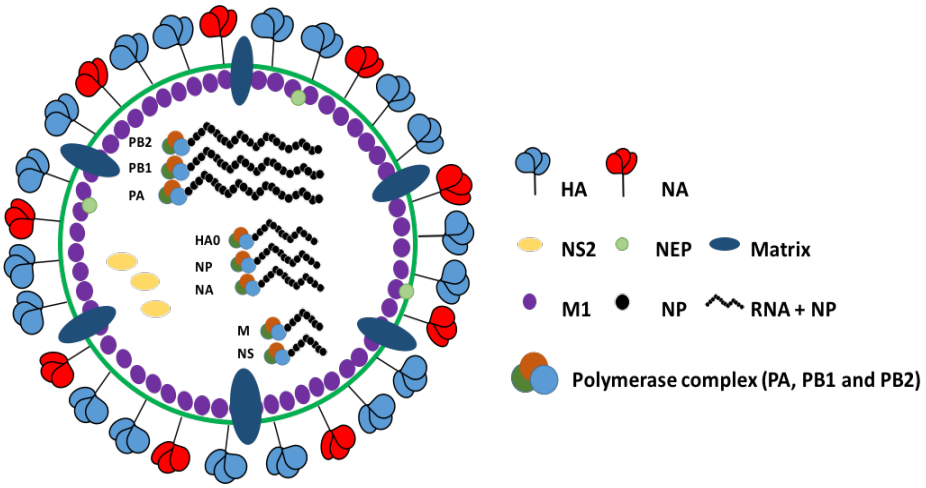


Figure 1. 1 The schematic structure of IAV. A single viral particle contains a lipid and enveloped layer. HA (blue spike) and NA (red spike) are embedded into the layer surface. There are eight genomic negative-sense single-stranded RNA segments in each IAV. IAV: Influenza A virus. M: Matrix protein. NP: nucleoprotein. NEP: nuclear export protein. PA: polymerase acidic protein. NS2: non-structural protein 2. PB1/2: polymerase basic protein 1/2.

Inside an IAV particle, there are eight segmented RNAs encoding 11 viral proteins: Haemagglutinin (HA), Neuraminidase (NA), Matrix proteins 1 and 2 (M1 and M2), Nucleoprotein (NP), Non-structural protein 2 (NS2), Non-structural protein 1 (NSP1), Polymerase acidic protein (PA), Polymerase basic proteins 1 and 2 (PB1 and PB2), and Polymerase basic protein 1 – F2 (PB1-F2) (Samji, 2009). Each viral RNA segment is associated with a polymerase complex (PA, PB1 and PB2), and multiple copies of NP, forming a viral ribonucleoprotein (vRNP) unit (Dawson et al., 2017) (**Figure 1.2**).

The viral glycoproteins HA and NA are responsible for the virus-host interaction: HA is for virus entering into host cells; and NA is for cleaving sialic acid (SA) receptors and releasing the viral progeny (Byrd-Leotis et al., 2017). To date, 18 18 different IAV HA molecules (H1-18), and 11 different IAV NA molecules (NA1-11) have been identified (Sleman, 2018). All the possible combinations of these two proteins are found circulating in wild aquatic birds and sporadically transmitting to

other species (Verhagen et al., 2017), except for H17 and H18, which have only been found in bats (Chavan et al., 2016).

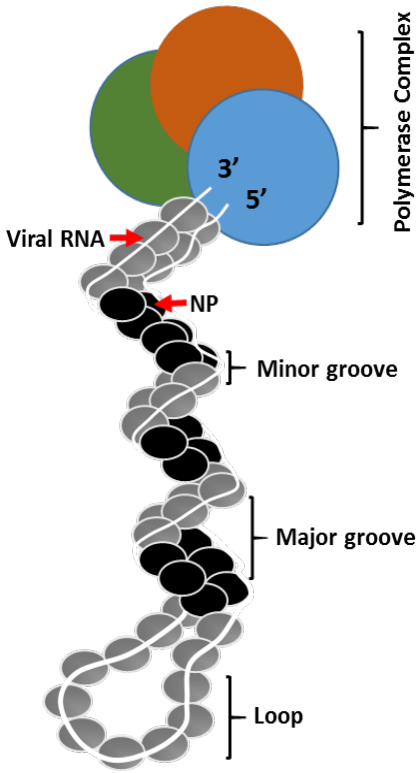


Figure 1. 2 One single vRNP complex.
NP encapsidates the viral RNA and forms vRNP together with other polymerases. In the helix, there are 'minor groove', 'major groove', and a loose region. The 5' and 3' vRNAs end at the binding sites of the polymerase complex.

1.1.3 Replication process

The interaction between the viral HA and the host SA-containing receptors is an important aspect of IAV infection along airway epithelium (Ibricevic et al., 2006). As shown in **Figure 1.3**, IAV is internalized into the cell cytosol by endocytosis (Edinger et al., 2014). In the host cells, the conformational changes of HA is triggered by the low pH of endosomes (endosomal acidification), leading to the fusion of viral and endosome surfaces (Hamilton et al., 2012). This allows a viral particle to incorporate into the cytoplasm through H^+ ions activated M2 channel. vRNPs are then released into the host nucleus through the endosome.

When entering into the nucleus, vRNPs are escorted by the importin proteins (Wu et al., 2007). In particular, the importin α can recognise the nuclear localisation signal (NLS) cargo and then binds to the importin β, which then mediates the translocation of the complex into the nucleus through the nuclear pore complexes (NPCs) (Wente and Rout, 2010). Once in nucleus, the negative-stranded viral RNA is converted to the positive-sense RNA as a template. Since a viral messenger RNA (mRNA) has a 3' poly(A) tail but not a 5' cap, viral PA cleaves off the 5' methylated cap of host cellular messenger RNAs (mRNAs) through a cap-snatching mechanism. The cRNA can act as a template for the production of negative-sense vRNA (Samji, 2009). New viral ribonucleoprotein (vRNPs) are imported into the nucleus and form new complementary RNPs (cRNPs) by associating with complementary RNAs. These matured vRNPs are trafficked to the cell membrane for packaging and budding, and releasing to infect neighbouring cells (Nayak et al., 2009).

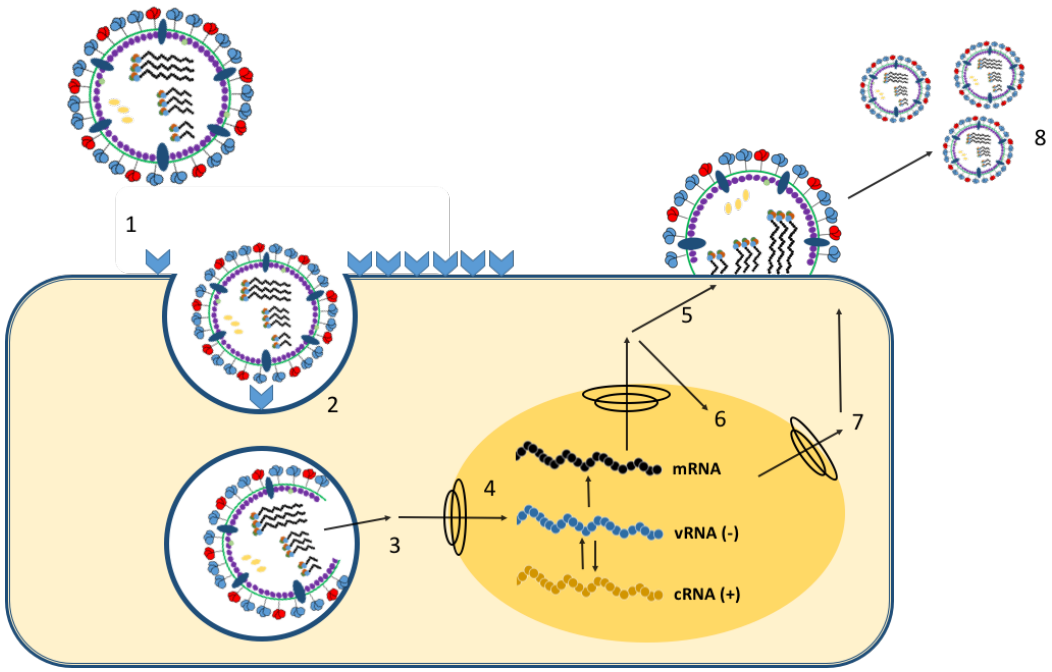


Figure 1. 3 Schematic diagram of the IV life cycle. Attachment: IAV life cycle is initiated when viral HA binds to SA-containing cell receptors along the airway epithelium (1). This is followed by the endocytosis into the cell (2). Fusion and uncoating: In an acidified environment, HA mediates the fusion of viral and endosomal membranes, which is followed by viral materials being released into the cytoplasm (3). Transcription and replication: Then, vRNPs is transported into the nucleus (4), where the transcription and the replication begin (5) and (6). Assembly: viral RNAs are assembled and matured, and transported to the plasma membrane (7). Budding: the buddings of these progeny virions are subsequently cleaved by NA, and are released from the infected cells (8).

1.2 Viral-host interaction

The virus-host interactions of HA with SA-containing receptors on host cell surfaces are critical in initiating viral entry, involving viral attachment, uptake, trafficking, and fusion (Skehel and Wiley, 2003). HA is to SA as a 'key' (**Section 1.2.1**) is to a 'lock' (**Section 1.2.2**).

1.2.1 'Key'—viral haemagglutinin

The haemagglutinin (HA), as a homotrimer, is accounting for more than 70% of the spike glycoproteins on the viral lipid membrane (Wilks et al., 2012). HA is consisted of two structurally distinct regions: a stem/stalk region and a head region. The stem region has a triple-stranded coil of α-helics, and the head region contains an antiparallel beta-sheet (on the top of the stem). The HA precursor, HA_0, can be cleaved by host proteases into two subunits in the lung: HA1 and HA2 polypeptides, which are responsible for receptor binding and mediating membrane fusion, respectively (Byrd-Leotis et al., 2017). More precisely, HA1 harbours receptor binding sites (RBS) and preferentially interacts with the host SA receptor and antigenic sites (Lazniewski et al., 2018); HA2 mediates a conformational change that causes the fusion of the viral and host membranes, allowing the IAV genome to enter into the cell via ion channels (Hamilton et al., 2012).

In summary, HA is responsible for the binding of viral particles to SA-containing receptors on the hosts, as well as involved in the fusion into the host cells (Xu et al., 2010). The alterations of HA are critical in host adaptation and cross-species transmission for multiple IAVs (Armstrong et al., 2000).

1.2.2 'Lock'—host receptors

The cross-species transmission of IAVs partially relies on the binding affinity of viral HA to specific sialylated glycans receptors distributed and exposed on exterior surfaces of cells (Mair et al., 2014). Surface sialylated glycans are covalently linked glycoconjugates on the cell surface, more precisely, sugar moieties attach to glycoproteins and/or glycolipids via asparagine residues (N-linked glycans) or serine/threonine residues (O-linked glycans) (Vasudevan and Haltiwanger, 2014). Sialylated N-linked glycans play a critical role in protein folding and stability, and cell-cell/pathogen interaction (Iijima et al., 2007), particularly it is essential for IAV entry (Chu and Whittaker, 2004).

Sialic acids (SAs) are electro-negative charged nine-carbon sugar. They are widely distributed on the outmost termini of N- or O-linked glycans catalysed by the sialyltransferase (ST) family on the cell membranes of mammals and some lower vertebrates (Varki, 2008). There are two frequent occurring forms of SA: N-acetyl-neuraminic acid (Neu5Ac), which is ubiquitous and most abundant form in all species; and 2-keto-3-deoxynononic acid (KDN), which is primarily found in prokaryotes and plants (Watson et al., 2015). N-glycolyl-neuraminic acid (Neu5Gc) is a derivative of Neu5Ac. It is formed when the amino group of Neu is replaced by an acetyl or glycolyl moiety (Mwangi and Bansal, 2003), and thus they are different in a single oxygen atom (**Figure 1.4**). Neu5Gc has been reported to be prevailed in pigs, but not generally in normal adult tissues of humans (Sriwilaijaroen et al., 2011).

SA is attached to the penultimate galactose (Gal) residues in α 2,3-linkage or α 2,6-linkage, and these two linkages determine the *trans*- or *cis*- conformation is adopted by HA of avian IAVs or human IAVs, respectively **(Figure 1. 5)**.

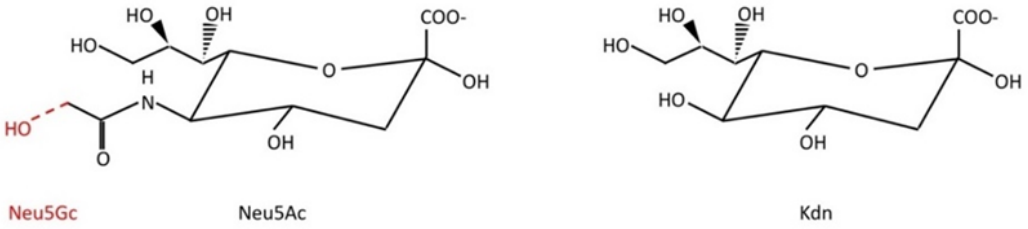


Figure 1. 4 Structures of Neu5Ac, Neu5Gc and KDN. Neu5Ac and Neu5Gc differ by an oxygen atom (left panel). It is added to CMP-Neu5Ac by cytidine monophosphate Neu5Ac hydroxylase (CMAH).

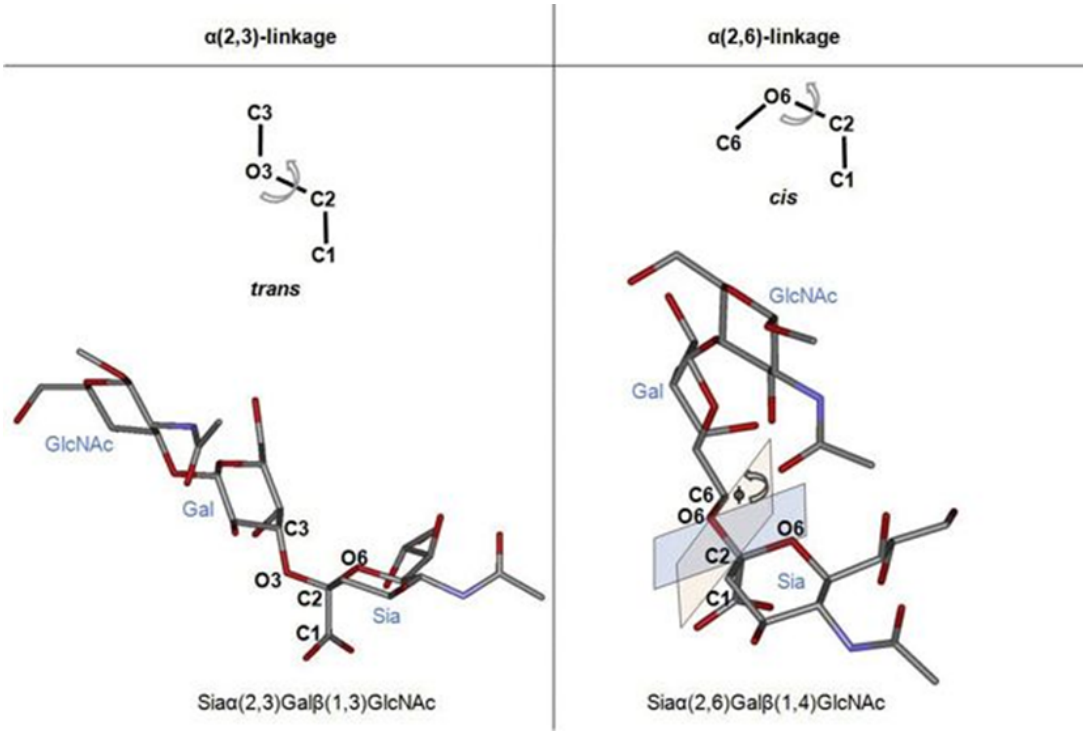


Figure 1. 5 Conformational structures of SA α 2,3-Gal and SA α 2,6-Gal. The left panel shows the SA α 2,3-Gal and the right panel shows the SA α 2,6-linked Gal. The conformational structure of SA has an impact on the affinity of HA binding (Datta and Paulson, 1997). Avian IAV prefers short and cone-shaped α 2,3-linked SA molecules and they have higher conformational flexibility (Wilks et al., 2012); Human IAV prefers long and umbrella-shaped α 2,6-linked SA. The image is originally from Jongkon (Jongkon et al., 2009).

As the first step of IAV infection, the interplay between HA and host SA-containing receptors is critical in IAV entry and efficient infection. Moreover, it is associated with IAV host switching, and viral interspecies adaptation.

1.2.3 Reservoirs, evolution and transmission

1.2.3.1 Host range and interspecies transmission

IAVs can infect a wide range of avian and mammalian species, including humans, marine mammals (seals and whales), swine, minks (ferrets), dogs, horses, bats, cows and cats (Parrish et al., 2015). IAV is mutated frequently in its HA and NA glycoproteins and display rapid evolutionary dynamics across different hosts, causing zoonotic infections and interspecies transmissions (Shao et al., 2017).

A comprehensive diagram of the interspecies transmission is shown in **Figure 1.6.** Taking avian IAVs as an example, such as *Anseriformes* (geese, duck, and swans) and *Charadriiformes* (gulls, terns, and waders), wild aquatic birds are natural reservoirs of IAVs. The replication of avian IAVs occurs along the epithelium of the respiratory tract and the lower intestinal tract. Among these IAVs, the high pathogenic avian influenza viruses (HPAIVs) and the low pathogenic avian influenza viruses (LPAIVs) differ by the basic amino acids at the cleavage site of HA_0 precursors. HPAIVs, such as H5N1, H7N7, and H7N9, are highly virulent, may cause influenza with high morbidity and mortality rates, which is also termed ‘fowl plague’ (Heider et al., 2015). Since 2003, HPAIV H5N1 has been reported to infect humans and transmit in species, posing a devastating impact on the agricultural production and public health (Alexander, 2006). HPAIV viral strains can result in a series of pro-inflammatory responses in a wide range of tissues, such as in brain, heart, liver and spleen. In contrast,

LPAIV only results in a mild infection along the respiratory and gastrointestinal tract of domestic poultries (Samy and Naguib, 2018).

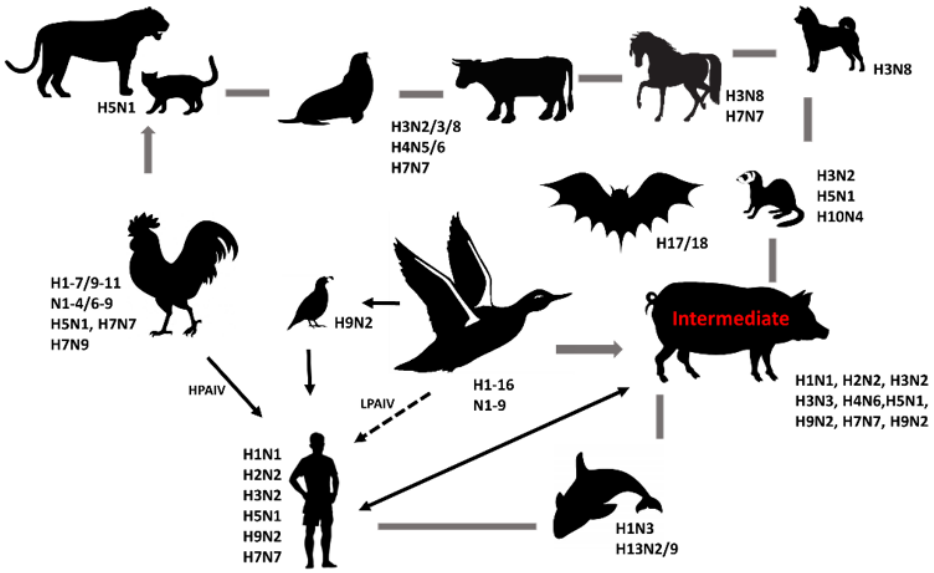


Figure 1. 6 Interspecies transmissions. In the interspecies transmission, pigs play intermediate roles in the influenza viral transmission. In specific, H5N1, H7N7, and H7N9, as high HPAIVs, can infect humans, but not transmitting or causing pandemics. Solid and bold arrows are the direct IAV transmission and dashed arrows shows limited infections. HPAIV: high pathogenic avian influenza virus; LPAIV: low pathogenic avian influenza virus.

1.2.3.2 IAV evolution by mutations and viral re-assortment

Human-adapted IAVs are able to escape human immunity pressures, causing seasonal influenza in human populations. When mutations take place in viral glycoproteins, 'antigenic drift' can occur as a gradual and subtle evolution of IAV. The mutated viruses can escape the recognition of host immune systems. Sometimes this may result in outbreaks of seasonal influenza with high morbidity and low mortality, known as 'epidemics' (Shao et al., 2017).

Pandemics may occur when annual epidemic events spread within human populations on a global scale (AB et al., 2018). Pandemics may outbreak in the naïve human population, and they are usually caused by 'antigenic shift'. New subtypes of viruses are introduced through a viral re-assortment of genome segments from more than two distinct influenza strains (Van Poucke, 2013).

1.2.3.2.1 Epidemics

The seasonal epidemics of human Influenza cause 290,000 to 650,000 deaths every year globally (WHO, 2018). The occurrence of epidemics are closely associated with seasonal changes. In temperate climates (e.g., United States), the influenza season falls in the winter months, starting from late fall, and peaking in mid to late winter. This pattern mirrors in the temperate areas of the southern hemisphere. In tropical regions, seasonal variations of influenza are less evident, and periodic milder epidemics occur throughout the year (Tamerius et al., 2011).

1.2.3.2.2 Pandemics

There have been five confirmed IAV pandemics since the twentieth century. Spanish 1918/H1N1, Asian 1957/H2N2, Hong Kong 1968/H3N2, Russian 1977/H1N1 and Swine 2009/H1N1. All these human pandemics occurred were related with the introductions of avian or swine IAV to human populations.

Hosts can be simultaneously infected by more than one viral strain, and viral genome from different strains can be mixed in one animal vessel. For example, this happened in the cases of 'Asian 1957/H2N2' and 'Hong Kong 1968/H3N2'. Moreover, viruses from non-human hosts can cross the species barriers and infect humans, as in the cases of 'Spanish 1918/H1N1' and 'Swine 2009/H1N1' (Taubenberger and Morens, 2013). Specifically, 'Spanish 1918/H1N1' was the first devastating pandemic outbreak in the 20th century. It was spread from North America to Europe by soldiers during the World War I, resulting in an enormous economic burden and approximately 50 million deaths globally in 1918. Firstly, 'Spanish 1918/H1N1' was considered a result of a genetic assortment between human, avian and swine IAV strains. However, Worobey later argued that it was a complete avian virus adapted to humans (Worobey et al., 2014). Although the accurate origin of the 1918 pandemic IAV still remains vague, all subsequent human pandemic IAV strains have been reported to be its descendants (Salomon and Webster, 2009).

1.2.3.2.3 The roles of pigs in IAV cross-species transmission

Swine H1N1 was the first IAV isolated from a pig (Shope, 1931). In 1976, the first case of a swine-to-human transmission of H1N1 was identified, although there was no pandemic immediately afterwards. It was not until 2009 that the Swine H1N1

influenza (pdmH1N1) took place in Mexico and spread to the rest of the world, killing almost half a million people. pdmH1N1 was reported to be a novel reassortant virus consisting of RNA segments from multiple sources: its PB1, PB2, PA, NP and NS genes were from a triple-reassortant North American H3N2 swine virus; its HA gene was from a classical swine H1N1 virus; its NA and M genes were from Eurasian avian-like swine lineage (Garten et al., 2009).

Therefore, although there is a major barrier in IAV interspecies transmissions, swine could act as an 'intermediate host' and a 'mixing vessel' for the generation of mutant IAVs, which can circulate in the human population with epidemic and pandemic potentials (Ma et al., 2009). It is thus of great practical relevance to come up with effective measures to prevent transmissions of IAVs to/from pigs and to reduce the susceptibility of pigs to IAVs.

1.3 Prevention and resistance

The surveillance for IAVs in pig herds has not been fully developed, and this hindered our understanding of the emergence of novel IAV strains, the transmission in pig breeding herds and to human populations.

1.3.1 Vaccines and drugs

The identified IAV epitopes provide sources for the production of vaccination and the discovery of novel antiviral drugs.

To date, vaccination against influenza has been applied as a cost-effective method to prevent influenza infection. In particular, live attenuated influenza vaccine (LAIV) strains have been proved to be safe, and genetically stable with a high level of efficacy against a broad range of influenza viruses (Jin and Chen, 2014). These cold-adapted, temperature-sensitive, and attenuated vaccine strains are produced by reverse genetics using HA and NA genes from the viral strains recommended by WHO (Bridges et al., 2003). LAIV is, for example, delivered in an intranasal manner to stimulate a long-lasting host humoral and cellular immunity (Jang and Seong, 2012). However, LAIV is contraindicated in humans that are immunocompromised, and is less effective against avian zoonotic viral strains because they do not replicate along the human upper airway (Shinya et al., 2006). Inactivated influenza vaccine (IIV) is a subunit vaccine containing purified HA and is administered intramuscularly (Grohskopf et al., 2014). It is commonly used to combat antigenically closely related influenza strains. However, IIV displays low immunogenicity and requires high antigen doses.

Currently, antiviral drugs offer an alterntaive route against IAVs but achieve only limited success. Existing antiviral drugs are consisting of two predominant classes: one is NA inhibitors, which inhibit the release of the viral progeny. Currently, the only licensed NA inhibitors, Zanamivir (Ralenza) and Oseltamivir (Tamiflu), have been approved by FDA (Englund, 2002). The other one is M2 inhibitors, which prevent the fusion of the virus and the immigration of the vRNPs complex into the nucleus. Amantadine and Rimantadine are the two licensed M2 inhibitors developed (Boltz et al., 2010).

However, the process of developing effective vaccinations and antiviral drugs for the novel viruses can be lengthy. Therefore, it is essential to study and develop alternative strategies that are universal and more effective against the ever-changing influenza

1.3.2 Host genetic control

To date there is little application of genetic strategies to mitigate IAVs. There has been considerable genetic research around this disease but the outcomes have largely been limited to the bench rather than the field or clinic. Reflecting this, host determinants and genetic variations have been less studied than viral factors. Since 2009, WHO has identified the major role of host genetics in the susceptibility to avian and human IV infection (Ruiz-Hernandez et al., 2016). Thus, host genes variability for virus-host interactions is now seen as a priority by some, especially the key genes controlling sialylated glycan conformation, assisting viral activities and escaping host immune responses, and dominant genes regarded as the targets for vaccination.

1.3.2.1 Animal model selection

When conducting a genetic-control approach in the hosts, selecting suitable animal models is a priority. They should be susceptible to infection and present similar responses with humans regarding clinical signs, efficient replications, viral kinetics, and transmission patterns. Therefore, ferrets, mice, hamsters, guinea pigs, and pigs have been identified to be suitable for modelling IAV host resistance.

Ferrets are suitable pathogenesis and transmission models, and have been used in IAV transmissibility and adaptation research for many years. They are naturally sensitive to human IAVs without prior adaptation, and present similar clinical signs in the upper airway tract with those of humans after infection. In 1933, a lab ferret was fortuitously inoculated with human IAV and was able to spread the virus to the co-housed ferrets (Smith et al., 1933). Moreover, the 2009 pandemic influenza (swine-origin H1N1) was reported to be highly transmissible among ferrets (Munster et al.,

2009). However, ferrets have limited availability and economically cost, hard to manipulate, and space demanding. Compared to ferrets, mice are less expensive and easier to handle. They have similar symptoms and immune responses as humans once infected, and thus have been widely used for modelling infections of some IAVs. However, they have relatively low susceptible to human IAVs. Moreover, IAV transmission in adult mice can be inefficient due to the lack of viral shedding (Ortigoza et al., 2018). Hamsters have been used as an alternative small-animal model. They present pathogenic effects along the airway when infected by human IAVs without adaptations, and thus are suitable models for human IAV infection and transmission (Iwatsuki-Horimoto et al., 2018). Guinea pigs have been occasionally used in IAV research (Azoulay-Dupuis et al., 1984). Compared with humans, they have similar anatomical and physiological properties, and comparable IAV receptor distributions. However, guinea pigs display no clear clinical signs after IAV infection (Sun et al., 2010).

Pigs are a valuable animal model in IAV research. They are similar to humans in many aspects, including but not limited to the genetics, anatomy, physiology, social behaviour, and clinical signs after infection. The immune system of pigs is well studied, allowing us to better understand the immunological mechanisms of IAV infections and replications in the hosts (Rajao et al., 2011). More importantly, pigs are natural hosts, and have ample susceptibility to both human and avian IAV strains (Meurens et al., 2012). Also, monitoring IAVs circulating in pigs is useful for the detection of newly emerging IAV strains that have pandemic potentials since pigs can play the role of 'mixing vessels' in generating novel IAV mutants (Ma et al., 2009).

1.3.2.2 Genetic-control approach in pigs

Starting from the first transgenic (TG) pig in 1985 (Hammer et al., 1985a), there have been a few attempts to engineer pig genes for pathogen resistance, disease eradication, and food production. For example, the F18 receptor gene of pigs was successfully engineered against oedema disease and post-weaning diarrhoea (Petters et al., 1997). 'Pig 26', produced through the editing of the RELA locus, had a potential resisilence to Afican swine fever (Lillico et al., 2016). Another example is the CD163-precisely-engineered pig, which is resistant to the Porcine Reproductive and Respiratory Syndrome (PRRS) virus (Christine Burkard, 2017). This study provides a potential strategy for host genetic resistance by precisely engineering host genes asscoiated with virus-host interactions, without sacrificing pigs' multifaceted and immune functions. The introduction of disease-resistant animal models by genome engineering sets a milestone in animal production and breeding — most likely the PRRS-resistant pig project will be first. Genetic resistance now represents a realistic and alternative solution to our battle against infectious disease transmission, preventing the potential disastrous pandemics that IAV can cause humanity.

1.4 My approach

The final goal of this project is to establish a pig model resistant to IAVs. To this end, based on a pig trachea cell model, I disrupt the IAV-receptor interaction by precisely engineering the α 2,6-sialyltransferase (ST6Gal1) coding gene, ST6GAL1.

The ST6Gal1 protein catalyses the configuration of the SA α 2,6-Gal receptor and is thus important in the virus-host interaction. The SA α 2,6-Gal receptor is assembled when SA residues are transferred from a nucleotide mono-phospho-sugar donor, CMP-Neu5Ac, to the terminal position of N-linked glycans with α 2,6-linkage (Dall'Olio et al., 2000), and this process is mediated by the ST6Gal1 protein. Indeed, the close association between ST6Gal1 and viral entry has been confirmed by several reports. For example, Vero cells with ST6GAL1 overexpression are observed with higher levels of human IAV (H1N1 and H3N2) propagation (Li et al., 2011). Moreover, human airway cells with inhibited ST6GAL1 expression present reduced virus binding and internalisation (Wu et al., 2014). Inspired by this work, my aim is to apply the CRISPR/Cas9 system (detailed introduction is in **Section 4.1**) to create a pig trachea cell model deficient in ST6GAL1 expression, with a long-term goal to apply such a strategy in generating less susceptibility of pigs to IAVs in farmed pig herds.

However, ST6Gal1 is a suitable but a complex target. ST6Gal1 has multifaceted roles in a wide range of biological processes, such as immune responses (Hennet et al., 1998), circulatory half-life (Manhardt et al., 2017), and inflammation (Bohm et al., 2012). Also, much effort has been made to clarify the roles of ST6Gal1 in cell apoptosis (Anugraham et al., 2014, Antony et al., 2014), angiogenesis (Imamaki et al., 2018), cell metastatic behaviours (Schultz et al., 2016a), cancer stem cell maintenance (Schultz et al., 2016b), as well as tumour malignant progressions in

humans, including colon (Swindall and Bellis, 2011), liver (Harada et al., 1993), breast (O'Neill et al., 2004), cervix (Hockel et al., 1996), ovarian (Wichert et al., 2018) , pancreatic (Shyr et al., 2011), and choriocarcinoma (Rodrigues and Macauley, 2018). For example, ST6Gal1 deficient mice have normal physiology, morphology and behaviours, but present immunodeficiency, such as attenuated proliferation, reduced IgM levels in serum, impaired B cell responses (Hennet et al., 1998).

Driven by the concern that inactivating ST6GAL1 could confer deleterious phenotyping effects to animals, I have pursued a subtle strategy towards precisely engineering pig ST6GAL1 gene by altering its transcriptional profile in the respiratory tract. With this approach, I aim to deepen our understanding of the relation between ST6GAL1 gene and viral infection in pigs. Moreover, the strategy could be promising in creating genetically engineered pigs that are less susceptible to IAVs.

I was enthused by the thought that my project could lead to a genetic-control or 'self-protective' pig model against the sneaking IAV at the 'door' of the airway. The hypothesized results are:

(1) The expressions of pig ST6GAL1 mRNA transcripts are in a tissue-dependent manner;

(2) The perturbation of ST6GAL1 gene expression will have an influence on IAV infection *in vitro*;

(3) The precisely manipulation of ST6GAL1 gene will not compromise immune responses and other physiological activities that SA α 2,6-Gal receptors participate in.

2 Materials and Methods

2.1 Cells and tissues

2.1.1 Cells

2.1.1.1 Cell strains and culture medium

There were four types of cell lines used in this study, a new-born porcine trachea (NPTr) cell line, a pig kidney (PK) -15 cell line, an Madin-Darby Canine Kidney (MDCK) cell line and MDCK cells with the cDNA of human 2,6-sialyltransferase 1 (MDCK-SIAT1) cell line.

(1) **NPTr cell line:** It is epithelial-like, non-carcinoma and non-transformed cells generated from a 2-day-old piglet trachea, and the serial culturing of the primary cells (Ferrari et al., 2003). It provides a wide microbial susceptibility spectrum to pathogens. The cell line was kindly provided by Prof. Laurence Tiley (University of Cambridge) and it was originally established at the Istituto Zooprofilattico Spefimentale in Brescia.

(2) **PK15 cell line** (ATCC® CCL-33TM): it is amenable to manipulate and thus has been well used for testing the cutting efficiency of CRISPR reagents.

(3) **MDCK** and **MDCK-SIAT1** cells are both provided by Prof. Paul Digard group at the Roslin Institute. They have been use to determine viral titre (Helen M. Wise, 2012).

There are four types of cell culture medium used in this study, complete medium, serum-free medium (SFM), conditional medium and freezing medium. NPTr, PK15, and MDCK cells were cultured in complete medium. MDCK-SIAT1 cells were cultured in complete medium supplementary with 2 mM Glutamine. All cells were

cultured at 37°C with a humidified and sterile atmosphere of 5% CO_2. Media should be pre-warmed in a 37°C water bath unless otherwise stated.

(1) **Complete medium** (culture medium): Dulbecco's Modified Eagel Medium (D-MEM, GlutaMAx and sodium pyruvate, Life Technologies, Cat# 31966-021) supplemented with 10% Fetal bovine serum (FBS, Cat#12657029, Gibco™), 100 U/ml penicillin, and 100 µg/ml streptomycin (10% pen/strep, stabilized solution, P4333, Life Technologies).

(2) **Serum-free medium (SFM)**: D-MEM supplemented with 10% pen/strep.

(3) **Conditional medium**: sterile filtrated from spent culture medium of cultured cells through a minisart syringe filter (Cat# SLHA033SS, Merck Millipore). It was aliquoted and frozen for storage at -20°C.

(4) **Freezing medium**: complete medium supplemented with 10% FBS plus 10% DMSO (Dimethyl sulfoxide).

2.1.1.1 Cell subculture and counting

All adherent cells were grown in cell culture treated vessels (e.g., flasks, petri dishes, or multi-well plates) with appropriate culture medium. When the cell confluence reached 80-90% of the surface area, the cells were sub-cultured. Cells were washed twice with Phosphate buffered saline (PBS, pH=7.2, Cat# 70013-016, Thermo Fisher Scientific), followed by dissociation from the surfaces of culture vessels using Tryple™ Express Enzyme (TE, 1x, Cat#12604013, ThermoFisher Scientific). Approximately 25 µl of TE per cm^3 in added to the culture vessel, followed by the incubation at 37°C for approximately 5 min. Once the cells were detached, the TE was inactivated by adding at least a double volume of fresh culture medium. The mixture

was then transferred into a 15 ml falcon tube, centrifuged at 300 x g for 5 min at room temperature and the supernatant was removed. The cells were then passaged by gently re-suspending the pellet in fresh culture medium before seeding on the new culture vessel. All cells were grown in a 37°C, and 5% CO_2 incubator, and were regularly inspected under a microscope to examine health status and contamination.

If necessary, cells were counted using a hemocytometer and a coverslip according to the protocol (Abcam) . Briefly, 20 µl of cell suspension was gently mixed with 20 µl of 0.4% Trypan blue solution (Cat# T8154, Sigma-Aldrich) in an Eppendorf tube. 20 µl of the Trypan blue-treated cell suspension was applied to a hemocytometer by the capillary action, and live cells were counted under a light microscope with a 10x objective. Therefore, the number of viable cells/ml were determined using the following calculation where the cell count was the average of the counts from the 4 sets of 16 squares: **Viable cells / ml = Cell count x Dilution factor x 10^4**

2.1.1.2 Cell growth and viability

The CellTiter-Glo® Luminescent cell viability assay (Cat# G7572, Promega) was used to detect cell viability. The assay was carried out with 10^4 cells seeded in white 96-well plates filled with 40 µl / well viral growth medium (VGM). At 0 h, 8 h, 24 h and 48 h points of culturing, 40 µl of CellTiter-Glo® regent was added and gently mixed. The cells lysis reaction was incubated for 10 min at room temperature with moderate shaking. The resulting bioluminescence was then measured and it was correlated with the levels of released adenosine triphosphate (ATP) providing a read out of cell viability.

2.1.1.3 Cell cryopreservation and recovery

To freeze cells, 10^6 cells were harvested and re-suspended in 1.5 ml freezing medium and transferred to 1.5 ml cryo-tube vials. The tubes were placed in a Mr. Frosty™ (Thermo Fisher Scientific) and transferred to -80°C. The Mr. Frosty™ is an iso-propanol freezing container that can slow the freezing rate < 1°C/min. After 24 h, the cells were transferred to a -150°C freezer on dry ice for long-term storage.

To recover cells, frozen cells were rapidly transferred to 37°C when it left -150°C. The thawed cells were diluted with culture medium followed by the centrifugation in a 15 ml centrifuge tube at 200 x g for 5 min. The supernatant was aspirated and the cell pellet was gently re-suspended in fresh culture medium before transferring to a culture vessel.

2.1.1.4 Transfection

Cells were transfected with DNA plasmids by Neon™ electroporation system (Life Technologies) and Lipofectamine® 2000 system (Life Technologies) according to the manufacturers' protocols.

2.1.1.5.1 Neon™ electroporation transfection

Neon™ electroporation system and 100 µl reaction tips were used to transfer plasmids into PK15 cells. All bubbles should be avoided strictly. On the day of transfection, pre-seeded PK15 cells were anticipated to reach 80% confluence in a 6-well plate. Cells were washed, digested, inactivated, and pelleted by the centrifugation at 200 x g for 5 min at room temperature (**Section 2.1.1.2**). The supernatant was removed and the cell pellet was re-suspended by gently pipetting up and down in 150 µl Resuspension Buffer (R). 1 µg each plasmid was added in each well and mixed well, followed by taken up into a compatible pipette tip. The pipette tip was inserted into the chamber with a holder containing 3 ml of Electrolyte Buffer (E). The optimised condition was 1200 v pulse voltage, 20 ms pulse width and one electrical pulse. Once the process was complete, the pipette tip was unplugged from the device. Cells were then transferred into a new 6-well plate pre-filled with SFM and incubated at 37°C for downstream experiments.

2.1.1.5.2 Lipofectamine® 2000 transfection

Lipofectamine® 2000 system was utilized to transfect plasmids and/or HDR template into NPTr cells. NPTr cells were washed, disassociated, counted and inoculated 5 x 10^5 cells in 6-well plates prior to transfection. On the day of transfection,

once approximately 80% confluence, cells were washed, the wells were filled with culture medium, and they were prepared for transfection. Briefly,15 µl of Lipofectamine® 2000 regent was diluted with 150 µl Opti-MEM Medium (provided), labelled as 'A'; 2 µg plasmid was diluted in 150 µl Opti-MEM Medium, labelled as 'B'. 'B' was added to 'A' at 1:1 ratio, and tubes were incubated at room temperature for 10 min. The plasmid-lipid complex was evenly dropwise on the cells and incubated at 37°C.

In both protocols of transfection mentioned above, negative and positive controls should be included; in which same amount of cells were transfected with PBS and 1 µg empty pSL66 vector, respectively. Successful transfection was indicated by the observation of the expression of green fluorescent protein (GFP+). Fluorescent cells at 24 and 48 hours post transfection (h.p.t) were visualized using a Zeiss microscope fitted with a digital camera with Image J software (3x upright Widefield Fluorescence Microscopes). All single panels should be in a same green gain and an adjustable blue gain. The fluorescent cells were then individually sorted into multiple 96-well plates by the Fluorescence-activated cell sorting (FACs) described below.

2.1.1.5 Fluorescence-activated cell sorting

At 48 h.p.t, cells were dissociated from the 6-well plates, and re-suspended with up to 300 µl PBS in sterile FACs glass tubes. FACs analysis was carried out using a BD FACS Aria IIIu 4 laser Cell Sorter in the Bioimaging centre (the Roslin Institute, UoE). Briefly, the cell suspension was converted into the stream of fluorescent and cells separated into individual cells through the vibration. The droplets then flow through an electrostatic orientation system allowing the GFP+ cells to be identified

and sorted individually into single wells of multiple 96-well plates pre-filled with 50% conditional medium and 50% culture medium. The remainder of the GFP+ cells were collected as 'a mixed pool'. The cell clone expansion was followed for approximately 20 days, during which period cells should be inspected weekly. Marking the wells in which cells are growing and refilling with culture medium. Confluent clones were transferred into new 96-well plates. Upon reaching 80%-90% confluence, transferred clones were divided into two 96-well plates. One was used for the further culture, and the other used for the quick extraction (QE) of gDNA, to allow the genotyping of the clones.

2.1.2 Tissues

2.1.2.1 Tissue collections and storage

Tissue samples were harvested from piglets and adult pigs culled for other purposes with approval from the Institute's Ethics Committee. Pigs were sedated with an intramuscular injection of ketamine (Vetalar) at a dose rate of 6 mg/kg, Azaperone (Stresnil) at a dose rate of 1mg/kg, followed by an overdose of anaesthetic agent (Pentoject).

Tissues of interest were harvested, including the right and lower lobe of lung, the cartilage of trachea, the stem bronchus, the lower lobe of the liver, a section of spleen containing red and white pulp, the apex of the heart, the duodenum, the wall of the stomach and the renal cortex of kidney (**Figure 2.1**). Each tissue sample was then dissected into multiple portions of ~100 mg and submerged in 0.5 ml pre-chilled RNAlater^{TM} Stabilization Solution (100ml, Cat# AM7020, Invitrogen) and snap-frozen on dry ice before long-term storage at -80°C.

2.1.2.2 Tissue manipulation

To dissect pig tissues, each sample was transferred from the RNase inhibitor solution (Thermo Fisher Scientific) onto a 10 cm^3 petri-dish on ice. Approximately 1ml chilled TRIzol™ reagent (Cat#15596026, Invitrogen) was added into each 100 mg frozen tissue. The mixture was mixed well and then transferred into a lysing matrix D tube, followed by the homogenization in a FastPrep FP120 instrument (Cat#116006500, MP Biomedicals) at a speed of 2 m/s for 20 s. 200 µl 1-Bromo-3-chloropropane (BCP, sublimed grade, 99.99% trace metals basis, Cat#699152, Sigma-Aldrich) was added in to dissociate the homogenate. After the incubation for 5 min at room temperature, the reaction was centrifuged at 12,000 x g for 5 min at 4°C. Homogenates were removed, and the phase from the upper aqueous phase was collected for RNA extraction and RNeasy Clean up (**Section 2.3.2.1)**.

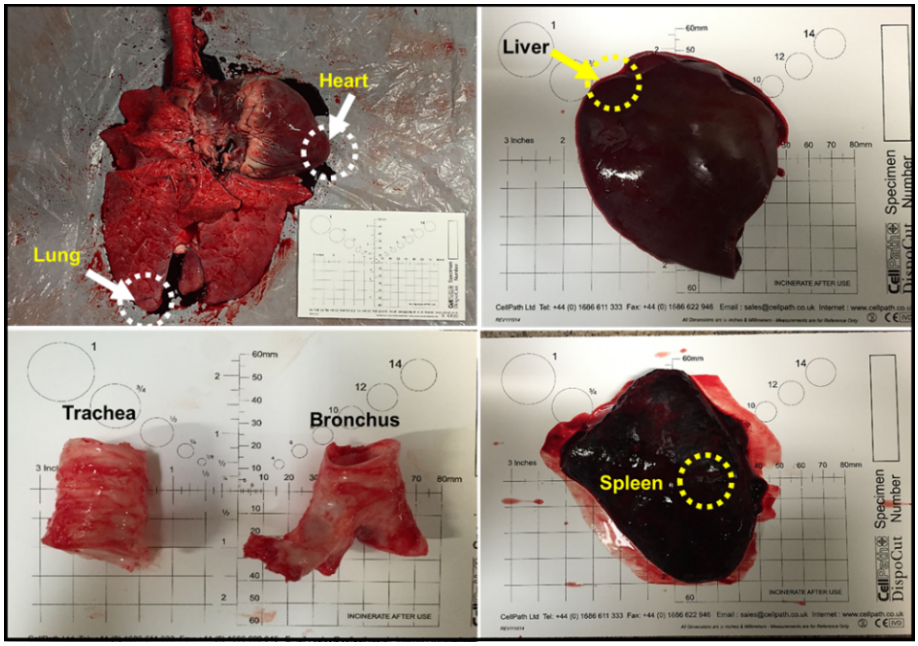


Figure 2. 1 Tissue extraction. Lung, trachea, bronchus, liver, spleen and heart samples collected are shown here. Dashed circles represent the portions isolated.

2.2 Microbiology techniques

2.2.1 Manipulation with *E. coli*

2.2.1.1 Agar Plates Preparation

Ampicillin (Cat# A0166, Sigma-Aldrich) was diluted to 50 mg/ml in Nuclease-free (NF) water. Luria broth (LB) liquid medium and agar were made by the Central Services Unit, the Roslin Institute. The final concentration of antibiotic-containing medium/agar was 100 μg/ml. The melted antibiotic-containing LB agar was evenly distributed on 90 mm single vent Petri dishes, which were then wrapped with cling film for storage at 4°C up to two weeks.

2.2.1.2 Transformation

Top10 competent cells (Life Technologies) were used in transformations. They were kept at -80°C before thawing on ice for transformation. 1-10 ng DNA (2 μl) was added in a vial of competent cells (50 μl). The reaction was mixed gently, and incubated on ice for 30 min. The tube had a heat shock at 42°C for 30 s, and was transferred on ice for further 2 min.

2.2.1.3 Cell plating and colony picking

Each transformation reaction was spread onto an antibiotic-containing LB agar plates. The plates were incubated overnight at 37 °C. The following day, a number of discrete colonies were picked up by a sterile tip and dip it into a colony PCR reaction for amplifications or kept in an antibiotic-containing LB liquid at 4°C for plasmid isolation.

2.2.2 Virology

2.2.2.1 Viral strains and culture medium

All experiments with IAV viruses were conducted under the biosecurity level three at the Roslin Institute. A lab-adapted strain (PR8: A/Puerto Rico/8/34), an H3N2 swine isolate (Swine87: A/Swine/England/163266/1987), and a mouse-adapted derivative of the prototype (Cal04: A/California/04/2009) were generated by reverse genetics using a previously described an eight-plasmid system (de Wit et al., 2004). Virus titres were determined according to the standard procedures described previously (Klimov et al., 2012)

There were two viral culture medium used in this study. **Overlay medium**: culture medium containing 0.14% (w/v) BSA, and this was used for plaque assay. **Viral growth medium**: culture medium containing 0.14% (w/v) BSA and 1 μg/ml TPCK-treated trypsin, and this was used for viral growth.

2.2.2.2 Viral infection

Cells were seeded on 24-well plates and cultured to achieve 80-90% confluence prior to viral infection. The culture medium was removed from cells. Cells were rinsed with SFM. Viral stock was quickly thawed at 37°C and diluted in an appropriate volume of VGM. The diluted virus was then inoculated in cells to provide a desirable multiplicity of infection (M.O.I), followed by the incubation at 37°C for one hour. Note that a negative control was included: one was not inoculated with virus.

For a single-cycle infection, any non-internalized and inactivate absorbed virus were removed from cells by aspirating the inoculum. Then cells were then overlaid with overlay medium. The NP expression was characterised by western blot and

immunofluorescence (IF) analysis. For a multiple-cycle infection, cells were overlaid with VGM directly and cultured at 37°C. The supernatant was harvested at desired time points post infection. The samples were snap-frozen on dry ice, and stored at -80°C for plaque assay (**Section 2.2.2.3**).

2.2.2.3 Plaque Assay

Growth kinetics in naïve and edited cells were performed using the three selected virus strains. Pre-seeded cells were washed with PBS and infection at an M.O.I of 0.001. Following absorption, cells were overlaid with VGM and samples were collected at 8, 24, and 48 hours post infection. All infected supernatants were collected and snap-frozen for being titrated plaque assay or storage at -80°C freezer.

Regarding virus titres, they were determined according to the standard procedures described previously (Klimov et al., 2012). The plaque assay for PR8 and Swine87 viruses was performed using MDCK cells. The titration for PR8 or Swine87 virus was performed using MDCK cells and for Cal04 virus was performed using MDCK-SIAT1 cells, allowing for individual plaques being readily discerned and counted after staining. Briefly, samples containing virus (**Section 2.2.2.2**) were ten-fold serially diluted in SFM. 400 µl of each of the dilution series was inoculated into each well and incubated at 37°C for 1 h, with an agitation every 20 min. After 1 h, the cell monolayers were overlaid with 2 ml overlay medium, supplemented with 1.2% (w/v) Avicel. Cells were then maintained in a 37°C incubator for 48 h.

For fixing and staining, the overlay medium was aspirated and cells were incubated with 2 ml of 10% neutral buffered formalin (NBF) at room temperature for 20 min. After that, NBF was removed, cells were covered with 5% Toluidine blue stain

and left for 30 min or overnight. Plaques were then counted in the well in which more than 10 plaques presented. The viral titre was determined according to the number of plaque forming units (PFU) per ml:

PFU/ml= Number of plaques/(dilution factor x volume of diluted virus).

2.2.2.4 Immunofluorescence

2.2.2.4.1 Immunofluorescent assay

An efficient infection is confirmed by the immunofluorescence assay (IFA) of intracellular nucleoprotein (NP), which is a RNA-binding protein pivotal to the early IAV replication (Portela and Digard, 2002). Following the single-cycle infection (**Section 2.2.2.2**), cells on the coverslips were washed gently, and fixed using 200 µl PBS/4%(v/v) formaldehyde for 15 min. The negative control was the one not inoculated with virus. After the fixing, samples were washed in PBS and permeabilized in 200 µl 5% PBS/0.2%(v/v) Triton-100x for 10 min. Followed by three washes with PBS, each sample was blocked in 200 µl PBS/10% FBS for 60 min. For the primary antibody incubation, each well was incubated with 200 µl primary antibody–contained PBS/1% FBS (Anti-IAV-nucleoprotein monoclonal antibody [AA5H], 1:1000 dilution, Cat#ab20343, Abcam) for 1 h at room temperature or overnight in a cold room. Cells were then washed with PBS/1% FBS for three times. For the secondary antibody incubation, cells were then incubated for 1 h, with 200 µl secondary antibody-contained PBS/1% FBS (Goat Anti-Mouse IgG H&L Alexa Fluor® 488, 1:1000 dilution, Cat#ab150113, Abcam) and DAPI (1:10,000 dilution, Invitrogen). Subsequently, cells were washed three times in PBS/1% FBS. Coverslips were mounted on a microscope slide with ProLong Gold Antifade reagent without DAPI

(Cat# P36934, ThermoFish Scientific). All the mounted slides were covered with foil forms and kept overnight at 4°C. On the second day, the fluorescent cells were visualised on a Leica DMLB-upright Fluorescent Microscope.

2.2.2.4.2 Flow cytometry analysis

In this step, all centrifugations were performed at 400 x g for 5 min at room temperature, and all washing steps were repeated three times unless otherwise stated. After a single-cycle infection (**Section 2.2.2.2**), cells were rinsed, lifted, and transferred into deep round bottom 96-well plates (P-96-450R-C-5PK, Kinesis Inc.). The plates were then centrifuged. Cell pellets were fixed with PBS/4% formaldehyde for 15 min, followed by washing with PBS/0.2% Tween-20. Cells were then permeabilised in 5% PBS/0.2%(v/v) Triton-100x for 10 min, followed by washing with PBS/0.2% Tween 20.

NP staining was performed essentially as described above (**Section 2.2.2.4.1**). The differences were as follows: the washing buffer was PBS/0.2% Tween 20 and the blocking buffer was PBS/2% BSA/0.2% Tween 20. Moreover, the primary antibody was diluted in 1:200 in PBS/0.2% Tween 20. The diluted primary antibody was added into cells and the reaction was incubated for 20 min at room temperature. Then, samples were be kept on ice and in the dark. Secondary antibody was in 1:700 dilutions in PBS/0.2% Tween 20, and added into the cells for a 15-min incubated. The final wash was in cold PBS/1% BSA buffer, followed by transferring cells into a polypropylene FCM tube. Data acquisition was performed using a BD LSR Fortessa X20 (14 colour Analyser) imaging flow cytometer with the filer B530-30A for Alexa fluor^{TM} 488 analysis. The data analysis was performed using Flowjo_V10 workspace.

2.2.2.4.3 Cytometry data analysis using FlowJo V10

A representative example of dot plots and the sequential gating strategy are described as follows: Forward-light scattering (FSC) and side-light scattering (SSC) were used to exclude the non-viable cells and cellular debris. After that, FSC-area (FSC-A) and FSC-height (FSC-H) were plotted to discriminate sticky or aggregated cell, and single cell events using a dot pot. Subsequently, a single-channel histogram plot was obtained, followed by the mathematical normalization. Therefore, the median fluorescent intensity (MFI) of each sample was measurement. A negative control was included to record the number of viable cell populations and the background signals.

2.3 Molecular biology techniques

2.3.1 DNA techniques

Please note that in this section, all procedures were performed at room temperature unless otherwise stated.

2.3.1.1 Plasmid DNA extraction

2.3.1.1.1 Small-scale of plasmid DNA (mini-prep)

Small-scale plasmid DNA extraction was performed using the PureLink™ Plasmid Miniprep Kit (Cat#K2100-11, Life technologies) according to the standard protocol. Briefly, overnight bacteria culture was pelleted at 4,000 x g for 5 min, and the supernatant removed. Pellets were re-suspended in 250 µl of RNaseA (20 mg/ml)-containing resuspension buffer (R3). 250 µl of lysis buffer (L7) was added to the cell pellet and the tube was gently inverted five times until the solution colour was uniformly blue. The lysis reaction was then incubated for 5 min, before adding 350 µl of Precipitation buffer (N4) to neutralise the reaction. The tube was then inverted gently until the solution became homogenous, followed by a centrifugation at 12,000 x g for 10 min to separate the cell debris. To bind the plasmid DNA on the silica spin column, the supernatant was centrifuged through a filter cartridge at 12,000 x g for 1 min, and the flow-through liquid was discarded. The bound DNA was washed three times with 700 µl ethanol-containing wash buffer (W9), and centrifuged again at 12,000 x g for 1 min. The DNA was eluted by adding 50 µl TE buffer to the centre of the column, and the flow-through DNA was collected in a clean recovery tube. The column was incubated for 1 min, before a final centrifugation at 12,000 x g for 1 min.

The concentration of plasmid DNA was quantified using a Nanodrop spectrophotometer (NanoDrop™ 1000, Thermo Scientific) (**Section 2.3.1.4**). Plasmid DNA was stored at -20°C for downstream experiments.

2.3.1.1.2 Large-scale of endotoxin-free plasmid DNA (Maxi-prep)

Large-scale plasmid DNA extraction was performed using Qiagen Endotoxin-Free (EndoFree) Plasmid kit (Cat#12362, Qiagen) according to the standard protocol. Briefly, 5 ml overnight bacteria culture was added into 100 ml antibiotic-containing LB medium, followed by the overnight culture at 37°C. On the following day, the overnight bacterial culture was centrifuged at 4,000 x g for 20 min at 4°C, and the supernatants removed. 10 ml RNase A and LyseBlue-containing resuspension buffer (P1) and was added in to re-suspend bacterial pellets by pipetting up and down. 10 ml lysis Buffer was then added in each reaction, and mixed by inverting the tube five times (no vortex), followed by the incubation for 5 min. Subsequently, 10 ml of pre-chilled neutralisation Buffer (P3) was added in the lysate, and the reaction was mixed thoroughly by gently inverting the tube for five times until the lysate became less viscous. The whole volume was transferred into a QIAfilter cartridge and incubated for 10 min. 50 ml Buffer ER was added into each filtered lysate and kept on ice for 30 min to prevent endotoxins binding. During the incubation, 10 ml of Buffer QBT was added to the QIAGEN-tip, and the column was equilibrated by the gravity flow. The lysate was transferred into the column, allowing it to flow through, and the reaction was washed with 30 ml Buffer QC for each. Then 15 ml Buffer QN was added on the centre of the column to elute the plasmid DNA into an endotoxin-free tube. The harvested DNA was then precipitated by adding 10 ml isopropanol, and the tube was

centrifuged at 4,000 x g for 60 min at 4°C. The supernatants were removed, and the DNA pellet was washed in 5 ml 70% endotoxin-free ethanol. The tubes were centrifuged at 4,000 x g for 30 min at 4°C, and then the ethanol was removed without disrupting the pellet. Subsequently, the tubes containing wet plasmid DNA were air dried in a clean hood for 5 - 10 min to evaporate ethanol residues. 100 µl of Elution Buffer (TE) was added to dissolve the DNA. The concentration of plasmid DNA was quantified using a Nanodrop spectrophotometer (NanoDrop™ 1000, Thermo Scientific) (**Section 2.3.1.4).** Plasmid DNA was stored at -20°C.

2.3.1.2 gDNA extraction

2.3.1.2.1 Small-scale gDNA extraction

Small-scale gDNA extraction was performed using DNeasy Blood & Tissue Kit (Cat# 69504, Qiagen) according to the standard protocol. Briefly, cell pellets harvested from 2-5 x 10^6 cultured cells were disrupted in 200 µl PBS, which contained 20 µl proteinase K. 200 µl Buffer AL was gently added to the reaction, and the reaction was incubated for 10 min in a 56°C water bath. 200 µl absolute ethanol was then added to the sample and mixed by pipetting. The mixture was transferred into a DNeasy Mini spin column, and the flow-through liquid was discarded after the centrifugation at 6,000 x g for 1 min. The column was inserted into a new collection tube, 500 µl Buffer AW1 was added, followed by the centrifugation and the removal of the flow-through liquid. Next, Buffer AW2 was added in the tube, followed by another centrifugation and a flow-through step. Finally, the column harbouring the gDNA of interest on the membrane was inserted into a clean 1.5 ml Eppendorf tube, and 100 µl of Elution buffer AE was added on the membrane of the column carefully. The tube was then

incubated for 1 min, and centrifuged at 12,000 x g for 1 min. A second elution was highly recommended to improve yields. The DNA can be directly used for PCR procedure (**Section 2.3.1.3**).

2.3.1.2.2 Quick extract gDNA from single-cell clones in 96-well formats

A rapid and efficient method of gDNA isolation was performed with QuickExtract™ extraction solution (QE, Cat#QE09050, Cambio). Well-grown cell clones were washed three times with PBS before adding 50 µl QE buffer. The reactions were mixed well, and transferred to skirted PCR 96-well plates. The plates were incubated at 65°C for 5 min, and 98°C for 2 min in a thermos-cycling PCR machine. The gDNA was then used for PCR amplifications (**Section 2.3.1.3**) or Sanger sequencing **(Section 2.3.1.8**). Alternatively, it can be kept at -20°C for long-term storage.

2.3.1.3 DNA amplification

PCR primers were designed using online tools Primer 3 Input (http://primer3.ut.ee/) and NCBI blast (https://blast.ncbi.nlm.nih.gov/Blast.cgi) unless otherwise stated. With some design criteria of primers for PCR, the general suggestion of GC content was ranged from 40 to 60%; annealing temperature was identified by a gradient PCR, with a range of temperatures setting from 50°C to 72°C; the optimum length ranged from 15 bp to 30 bp. After receiving the stock of lyophilised oligonucleotides, they were dissolved to 100 µM (stock concentration), and diluted to 10 µM (working concentration). All primers were kept at -20°C for long-term storage. All the primers for PCRs are listed below:

Primer	Sequence (5’ - 3’)
F2	TGTGGTCCTGGGCATTTCTC
R2	CCTTTGCACACCCTGACTCT
F3	TAGGAGCGACATCTGGGCTA
R3	GCCCTGTTGCACTTGCAAAG
GSP-1	GGACACTTTATACTTGTTC
GSP-2	AGCTGTCCTTGTTCCACACC
AAP	GGCCACGCGTCGACTAGTAC$(G)_{14}$
NP	GTGCCGCTGGAGGATGCAGA
UAP	$(CUA)_4$ GGCCAGGCGTCGACTAGTAC $(G)_{14}$
Validation of editing using PCR assays (below)	
F1	GACCAAGACAACTCGGAGGA
R7	TAGCCCCTGCGCTTTTTCTT
F2	GGAAACACAGAGGGAGGTCG
R2	GTCAGCCAGGAAGGAACCAA
F9	TGAATGGCTGGTTCGAGTCC
R9	ACTGCTGTAAGCCTCGTGTC
F6	TGAATGGCTGGTTCGAGTCC
R6	ACTGCTGTAAGCCTCGTGTC
E1F1	CCTGCTGGTCTTTCTCCTGT
E2R1	CACCCCAGAACCCACCTATT

Table 2. 1 Primers for DNA amplification. Detailed information of primers is described in the corresponding result chapters. Among these primers, F2 and R2, F3 and R3 were used to amply the region between coding Exon (1) and Exon (2); GSP-1, GSP-2, AAP, NP and UAP were used in rapid amplification of cDNA Ends (5’ RACE) experiment.

2.2.1.3.1 Routine PCR with Phusion DNA Polymerase

In this study, the reaction of routine PCRs were performed in a final volume of 25 µl. The PCR components included 12.5 µl Phusion® high-Fidelity PCR Master Mix (Cat#M0531S, NEB), 1.25 µl volume of 10 µM forward and reverse primer, 2 µl of 20 ng DNA and 8 µl NF water, and the final concentration of them, as well as the reaction protocol are shown in **Table 2. 3**.

Component	Final Concentration	
DNA template	20 - 100 ng	
Forward primer (10 µM)	0.5 µM	
Reverse primer (10 µM)	0.5 µM	
2x Phusion Master Mix	1 X	
Nuclease-Free (NF) water	Up to 25 µl	
Thermocycler conditions		
98°C	30s	
98°C 50°C-72°C 72°C	10s 20s 25s/kb	35 cycles
72°C	5min	
4°C	∞	

Table 2. 2 PCR conditions with Phusion DNA Polymerase. All the PCR reaction components were prepared on ice and quickly transferred to the preheated thermoycler machine (98°C). A negative control containing no templates was included.

If required, colony PCRs were performed with the supernatant of bacteria culture as templates. Briefly, after the plasmid transformation **(Section 2.2.1.2**), cell plating and overnight incubation of bacteria (**Section 2.2.1.3**), 1 µl of the supernatant of lysed bacterial was added in a 25 µl volume of PCR reaction containing the components introduced above (**Table 2.4**). A negative control without any transformation was included. The PCR reactions were performed with the conditions shown above, followed by the band visualization on 2% agarose gel (**Section 2.3.1.4**).

2.3.1.3.2 5'- RACE PCR with *Taq* polymerase

5' Rapid amplification of cDNA ends (5' RACE) PCR was performed using the kit (5' RACE System for Rapid Amplification of cDNA Ends, version 2.0, Cat# 18374058, ThermoFisher Scientific) (**Figure 2. 4**) and all the volume of components were added according to the standard protocol.

Briefly, prior to the 5' RACE PCR, the first strand of cDNA was synthesized using of a gene-specific primer 1 (GSP1) from total mRNA (**Section 2.3.2.1)**. After the first strand cDNA synthesis (**Section 2.3.2.2**), RNA and DNA heteroduplex molecules were removed by adding RNase Mix. Then the residues, containing the unincorporated dNTPs, GSP1, and proteins, were removed from the first strand of cDNA using S.N.A.P Column. Subsequently, terminal deoxynucleotidyl transferase (TdT) and Deoxycytidine triphosphate (dCTP) were added into the reaction to assemble a homopolymeric tail on the cDNA.

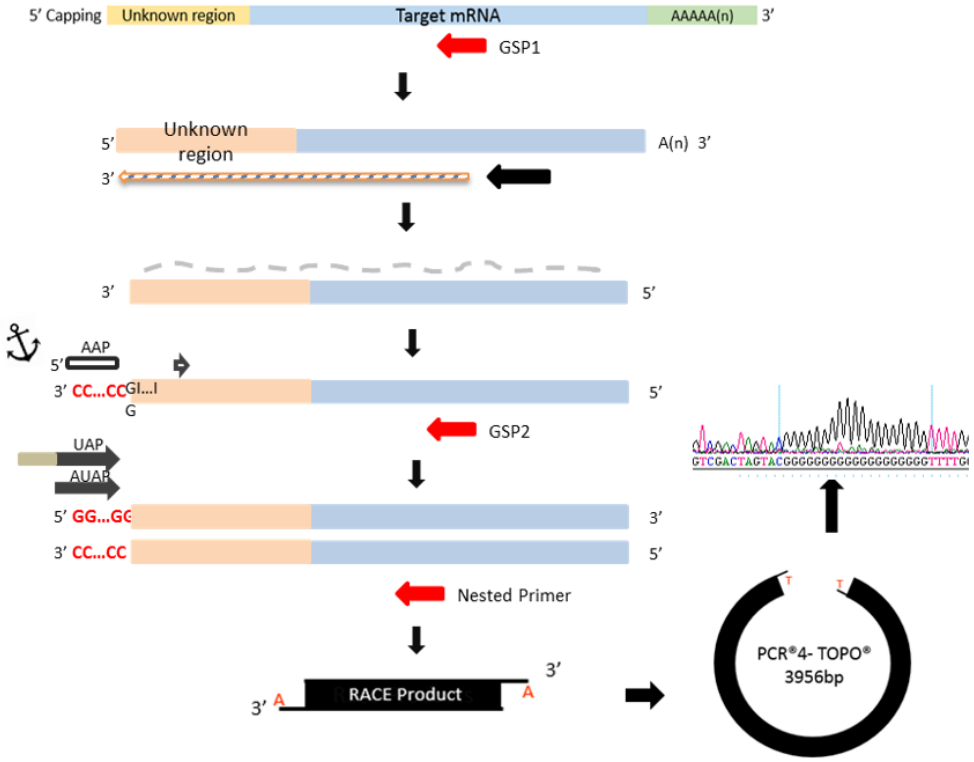


Figure 2. 2 Schematic overview of 5' RACE procedure. AUAP contains an adaptor region which was homologous to the adpator region of the AAP, and can be used to reamplify primary 5' RACE products. GSP-gene-specific primer; AAP-5' RACE abridged anchor primer; AUAP-abridged universal amplification primer.

After the tailing reaction, the first round of PCR was performed on the tailed cDNA with *Taq* DNA polymerase. The primers in this round of PCR were GSP2 and a deoxyinosine-containing anchor primer specificity of the PCR products, another round of PCR was performed with a nested Primer (NP) and Abridged Universal Amplification Primer (AUAP) using the first round of PCR product as a template (1:20 dilution). The NP was located upstream of GSP2 within the cDNA product. The reaction conditions of these two rounds of PCRs were optimised and described as below (**Table 2. 5**).

The second round of PCR product then visualised on a 2% agarose gel electrophoresis (Section 2.3.1.4). All the visualised bands were extracted, purified and subcloned into a pCR™4-TOPO® TA vector (Cat# 450030, ThermoFisher Scientific) for subsequent sequencing with M13 reverse primer (**Section 2.3.1.8)**.

Thermocycler conditions			
	PAD	94°C	1 min
35 cycles	Denaturation	94°C	1 min
	Annealing	55°C	1 min
	Extension	72°C	2 min
	Final Extension	72°C	6 min
		5°C	Pause

Table 2. 3 Optimized conditions of 5' RACE. The steps of denaturation, annealing and extension were cycled for 35 times.

2.3.1.4 DNA visualisation and quantification

UltraPure™ Agarose powder (Cat#17852, ThermoFisher Scientific) was added to 1 x TAE solution (40 mM Tris, 20 mM acetic acid, 1 mM EDTA; supplied by Central Services Unit, Roslin Institute) to make agarose gels with certain percentages. The mixture was boiled in a microwave for approximately 3 min to dissolve the agarose completely. 10% total volume of SYBR™ safe DNA Gel Stain (Cat# S33102, ThermoFisher Scientific) was added to the agarose solution just before pouring the solution into a casting tray.

Each restriction enzyme digested product or PCR product of interest was mixed with 6 x blue load dye (Cat# T8154, Sigma-Aldrich) (the final concentration was 1 x) and loaded into each well of a gel. GeneRuler 1 kb DNA ladder™ (Cat# SM0312, Thermo Scientific) was used as a size marker unless otherwise stated. Run the gel at 80-100 voltage (V) for 1 h until the marker was clearly distinguished so that the

approximate sizes of the products can be determined. Subsequently, the gels were visualised under a UV light released by an Ultraviolet (UV) Trans illuminator (256 nm).

DNA quantification was carried out using a Nanodrop ND-100 spectrometer (Nanodrop Technologies, Inc). After cleaning the sensor of the machine with water and a lint-free cloth, 1 µl of dilution buffer was loaded as a blank. Then 1 µl DNA sample was then loaded onto the clean sensor. The final concentration was reported in ng/µl.

2.3.1.5 DNA Purification

2.3.1.5.1 Gel bands isolation and purification

PCR products excised from agarose gels were purified using the QIAquick Gel Extraction Kit (50, Cat# 28704, Qiagen) according to manufacturer's instructions. Briefly, a gel slice was weighed in a 15 ml falcon tube before adding three volume of buffer QG of each gel weight (100 mg-100 µl). The sample was then incubated for 10 min in a 50°C water bath until the gel slice was dissolved completely. The tube was gently vortexed every 3 min to solubilize the gel. After that, 1 volume isopropanol was added to the mixture and the reaction was transferred to a QIAquick spin column inserted in a 2 ml collection tube. The column was centrifuged at 12,000 x g for 1 min and the flow-through liquid discarded. The column was washed by the addition of 750 µl Buffer PE followed by the centrifugation at 12,000 x g for 1 min. To elute the DNA, the column was transferred to a clean Eppendorf before loading 50 µl of elution buffer on the membrane of the column. The eluted DNA was quantified by the Nanodrop (**Section 2.3.1.4**) and stored at -20°C.

2.3.1.5.2 PCR products Purification

To clean up PCR products, such as removing primer dimers, extra dNTPs and enzymes, and failed PCR residues produced from PCR reactions, PureLink™ Quick PCR Purification Kit (Cat#k310001, ThermoFisher Scientific) was used according to manufacturer' s instructions. Briefly, four volumes of Binding Buffer (B2) was added to 50 µl PCR volume and the reaction was mixed well. The mixture was then transferred to a PureLink™ spin column inserted into a clean 1.5 ml collection tube, and centrifuged at 10,000 x g for 1 min. The flow-through liquid in the tube was discarded. 650 µl Wash Buffer mixed with ethanol was added to the column and centrifuged at 10,000 x g for 1 min. After this, the column was centrifuged at 12,000 x g for another 2 min to remove liquid residues. Finally, 20 - 50 µl Elution Buffer (E1) was added on the membrane of the column, and the reaction was incubated for 1 min, before the spin at 12,000 x g for 1 min. The elution step was performed twice to increase the yields of the clean-up PCR product.

2.3.1.6 DNA subcloning

PCR products amplified with *Taq* polymerase were subcloned into the pCR™4-TOPO®TA vectors using the TOPO™ TA Cloning™ Kit (Cat# 450030, Invitrogen) according to the standard protocol. **Figure 2. 4** illustrates the structure of the pCR™4-TOPO®TA vectors. The ligation reaction is detailed in **Table 2. 5**. All the components were vortexed and incubated at room temperature for 5 min, followed by the incubation on ice. Then the products were prepared for transformation (**Section 2.2.1.2**).

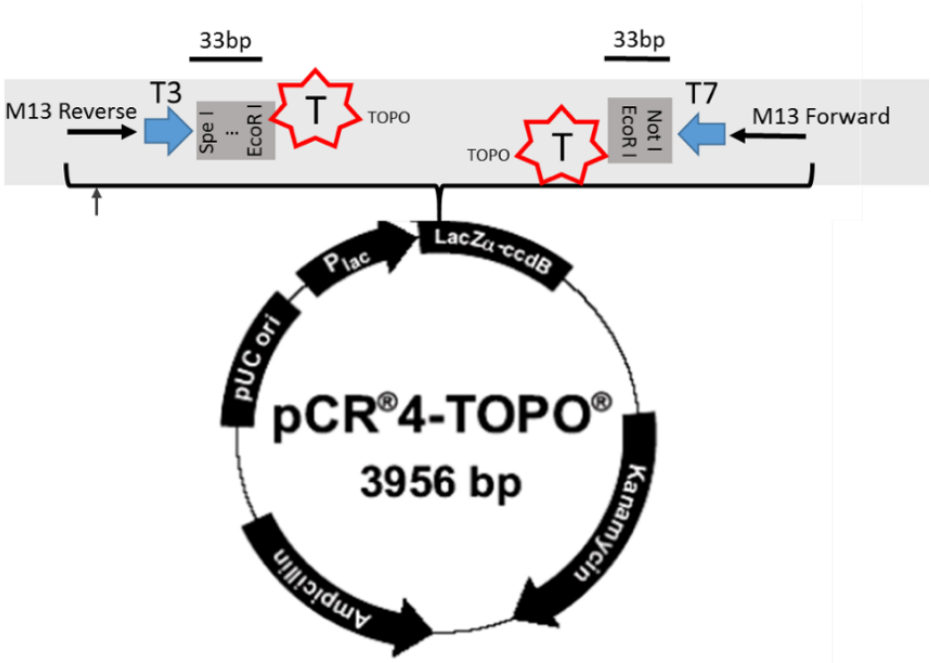


Figure 2. 3 pCR™ TOPO® TA cloning ® vector. The image was adapted from the manufacturer. M13: Forward/reverse priming sites; T7: promoter; Puc origin: LacZ.
M13 Reverse primer: 5′-GTAAAACGACGGCCAG-3’;
M13 Forward primer: 5′-CAGGAAACAGCTATGAC-3′

Compositions	Volume
pCR™4-TOPO®TA vector	1 µl
PCR product / gel purified product	10 ng
Salt solution (1.2 M)	1 µl
Sterile water	Up to 4 µl
Final Volume	6 µl

Table 2. 4 Ligation conditions.

2.3.1.7 DNA digestion

NEB supplied all restriction endonucleases unless otherwise stated. 20 units of enzyme (no more than 10% of total reaction volume) was applied in a proper reaction volume. The reaction was incubated at 37°C for 5 min for routine gDNA digestion, or overnight for HMW gDNA digestion **(Section 2.3.1.11).**

2.3.1.8 DNA Sequencing

DNA sequencing was performed by Edinburgh Genomics. PCR products were purified (**Section 2.3.1.3**) and then diluted to the final concentration according to the standard requirements. Briefly, 1.5-3 ng of DNA was required for a fragment of 100-200 bp product; 4.5-10 ng for 200 bp-500 bp product and 7.5-20ng for 500-1000 bp. Total 4 µl of PCR clean-up product and 1 µl of sequencing primer (6.4 pmol/µl) were mixed in a PCR-tube for sequencing.

Dirty PCR products in 96-well plates were sent to the in-house DNA sequencing facility (Sequencing Technical Services) at the IGMM (MRC Institute of Genetics & Molecular Medicine, University of Edinburgh, UK) for clean-up and sequencing. The PCR products (up to 10 ng/µl) were purified with their automated clean-up service via a liquid handling robot (Biomek ® NX Laboratory Automation

Workstation), in accordance with the manufacturer's instructions. Subsequently, Sanger sequencing reactions were performed utilising DNA Engine Tetrad ® 2 Peltier Thermal Cycler (BIORAD) PCR machines. The sequencing products were analysed on 3130xl (16 capillaries) and 3730 Genetic analysers (48 capillaries) (AB Applied Biosystems/Hitachi). All sequencing data was obtained as ABI files and analysed using Lasergene (DNASTAR series) and Geneious v1.2 Software.

2.3.1.9 Bioinformatics

The bioinformatics analysis was performed by Dr. Sylvia Beka. Firstly, the chromosomal location and sequence of pig ST6GAL1 were aligned to that in human ST6GAL1 gene using Ensembl genome browser, NCBI and the Blast software. Exonic and intronic sequences were predicted using Genscan (http://genes.mit.edu/GENSCAN.html) and Augustus (http://bioinf.uni-greifswald.de/augustus/) prediction software. Alignments were also performed with MegAlign (DNASTAR) software; and similarities of gene sequences between pigs and human were detected using the EBI Clustal-Omega tool (https://www.ebi.ac.uk/Tools/msa/clustalo/).

Promoter sequences were predicted using the promoter 2.0 prediction server (http://www.cbs.dtu.dk/services/Promoter/) and an online tool applied for the recognition of Pol II promoter sequences. Regulatory features were then predicated and identified by the level of highlights in the Ensembl browser. In parallel, evolutionary conservation of predicted pig promoters was based on 3 methods, (1) running multiple sequence alignments with predicted promoters using Clustal-Omega, (2) NCBI blast-blasting the predicted pig promoter sequences to look for significant

matches with other predicted mammalian promoters for same gene, (3) ECR browser searching for predicted promoters using the online tool ECRbase (https://ecrbase.dcode.org/). Finally, binding motifs in the pig promoters were identified using the Jasper database (http://jaspar.genereg.net/) and Transfac (http://gene-regulation.com/cgi-bin/pub/databases/transfac/search.cgi).

2.3.1.10 Assembly of CRISPR/Cas9 reagents

2.3.1.10.1 sgRNA design and synthesis

The web tool CCTop (https://crispr.cos.uni-heidelberg.de/) was used to aid in the predicting and ranking candidate CRISPR sgRNAs to the genome region of interest. A workflow shows a specific example demonstrating how this website works **(Figure 2.5)**.

Briefly, a 500 bp target sequence of pig ST6GAL1 gene was pasted into the CCTop software. All the parameters (e.g., protospacer adjacent motif (PAM) type, promoters and species) were input correctly. By clicking 'Submit', a list of eligible sgRNA candidates with on- and off-target scores were displayed. All evaluated sgRNA sequences were obtained and more detailed information of interest can be downloaded in an output as a *Fasta* file.

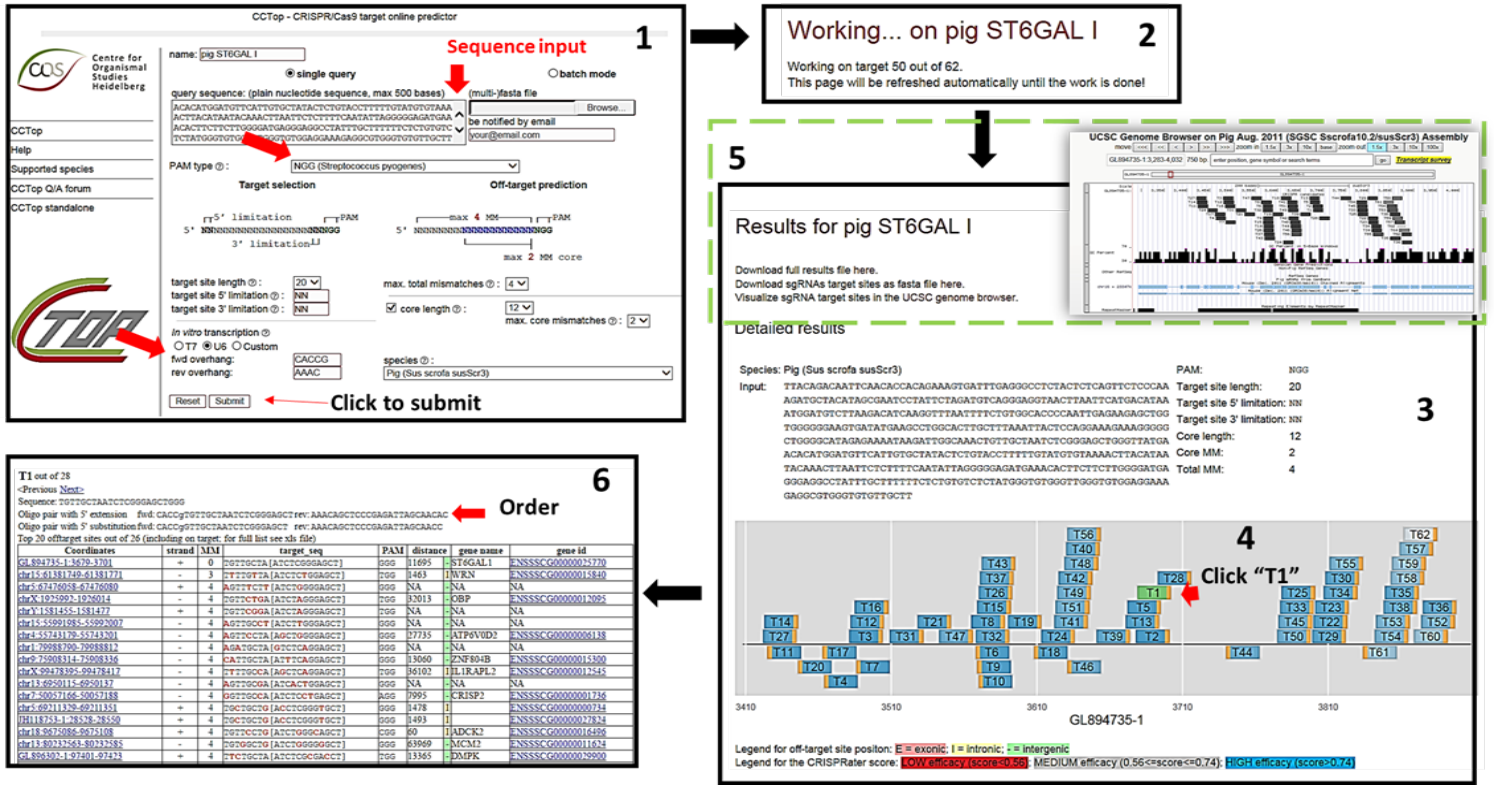


Figure 2. 4 CCTop online interface. In the screenshot, red arrows in (1) shows important input fields which are customizations for sgRNAs; (2) confirms the system is working normally; Consequently, from interface (3) and (4), sgRNA candidates are ranked according to their performance; More information of their genomic locations and GC abundance could be in further gained as shown in box 5; potential off-target effects caused by each sgRNA and their positions can be predicted, as shown in box 6. If the sgRNA selected meets the standards, the forward and reverse sequences are provided.

The sequences of all sgRNAs used in this study were detailed in the **Table 2.5**. Please note that the 5' of the sgRNAs should be a 'Guanine (G)' nucleotide, as it was required for the efficient transcription of human U6 RNA polymerase III (Bannister et al., 2007). Moreover, an appropriated single-stranded oligodeoxynucleotide (ssODN) template spanning the target site and complementary to the targeted sequence with homology arms, was designed as a donor template. The repair strategy of the donor template is schematically diagrammed in **Chapter 5**.

Primer	**Forward oligo (5' – 3')**	**Reverse oligo (3' – 5')**
SgRNA4	caccgCGAATCCTATTCTAGATGTC	aaacGACATCTAGAATAGGATTCGC
SgRNA5	caccgATTCGCTATGTAGCATCTT	aaacAAGATGCTACATAGCGAATC
SgRNA6	caccgTTTGGGAGAACTGAGAGTAG	aaacCTACTCTCAGTTCTCCCAAAC
SgRNA10	caccgCCACTTGATTGATAGCACGG	aaacCCGTGCTATCAATCAAGTGGC
SgRNA11	caccgATAGCACGGAGGTAATAGA	aaacTCTATTACCTCCGTGCTATC
SgRNA12	caccgCGGAGGTAATAGAGGGCTTT	aaacAAAGCCCTCTATTACCTCCGC
SgRNA4	caccgTCCCTTCAATACCTCTGAAT	aaacATTCAGAGGTATTGAAGGGA
SgRNA10	caccgTGTGCTGTCGTCTCATCAGC	aaacGCTGATGAGACGACAGCACA
SgRNA13	caccgTTCAATACCTCTGAATGGGA	aaacTCCCATTCAGAGGTATTGAA
HDR template	TAGAAGCCACAGATTTTCCCTTCAATACCTCTGAATTCTAGGGTTACCTGCCAAAGGAGAACATCAGGACCAAGGCCGGGCCGTGGGGCAGGTGTGCTGTCGTCTCATCAGCGGGATCTCTCAAGTCCTC	Introduction of a pre-mature stop codon by GGG (PAM)>TCT and an EcoRI restriction site

Table 2. 5 Oligos for CRISPR work. 'cacc' (red) is the 5' overhangs which were compatible with *BbsI* restriction enzyme sites in the plasmid containing Cas9.

All oligos used in this section were synthesised by Integrated DNA Technologies, Inc. (IDT). Synthesised oligo powder was re-suspended in TE buffer to reach a final concentration of 100 µm.

2.3.1.10.2 Assembly of CRISPR/Cas9 reagents

A cloning vector pSL66 containing *BbsI* restriction enzyme sites was used for the co-expression of the gRNA and Cas9 protein (**Appendix 7**). The pSL66 vector is a modified version of p*Sp*Cas9 (BB)-2A-eGFP (pX458) (Addgene plasmid # 48138) as described (Chen et al., 2013). The pSL66 vector contains a hU6 promoter driving the transcription of the gRNA; also, a hybrid cytomegalovirus (CMV) enhancer/chicken β-actin (CAG) promoter is involved in the pSL66 vector to enhance the expression levels of p*Sp*Cas9 (BB)-2A-eGFP in cells.

The annealed gRNA oligonucleotides were ligated into the pSL66 vector at the *BbsI* sites. All the procedures are listed below in **Table 2. 7**. After this, the reaction was transformed into competent bacterial DH5α cells (**Section 2.2.1.2**), and the correct insertions were confirmed by PCR and sequencing. The correct pSL66-gRNA expression construct was then transfected into PK 15 cells or NPTr cells (**Section 2.1.1.5)**, followed by the estimation of the cutting efficiency of CRISPR-Cas9 complexes by the T7 endonuclease I (T7EI) mismatch assay (**Section 2.3.1.10.3**).

(a)Annealing complementary oligonucleotides

Components	Volume (µL)
sgRNA sense oligo (100 µM)	0.5
sgRNA anti-sense oligo (100 µM)	0.5
10x NEB Buffer 2	0.5
NF water	3.5

Thermocycling conditions	
37°C	30 min
95°C	5 min
25°C	5°/min decreasing

(b)Insertion into pSL 66 vector

Components	Volume (µL)
pSL66 (1 µg/ µl)	0.5
Diluted oligo duplex	0.5
10x buffer	1
DTT (10 mM)	0.5
ATP (10 mM)	0.5
FastDigest BbsI (10 U/µl)	0.5
T7 ligase (2000 U/µl)	0.25
NF water	6.25

Thermocycling conditions		
37°C	5 min	6 cycles
21°C	5 min	

(c)Plasmid Safe

Components	Volume (µL)
Reaction from above	10 µL
10x Plasmid Safe Buffer	2 µL
ATP (10mM)	1 mM
PlasmidSafe exonuclease	1 µL
NF water	Up to 20 µL

Thermocycling conditions	
37°C	30 min
70°C	30 min

Table 2. 6 The conditions of CRISPR/Cas9 construct assembly. Complementary gRNA oligos are synthesised and re-suspended in TE buffer to reach a final concentration of 100 µm (a). Then annealed oligos from reaction is diluted with TE buffer, followed by the insertion into PsL66 vector at *BbsI* restriction digest sites (b). To remove undigested and failed ligated products, a plasmid Safe exonuclease treatment is suggested (c).

2.3.1.10.3 Assessment of cutting activity of CRISPR guides

T7E1 assay

A 300-700 bp length of DNA sequence flanking the target locus was amplified and purified before the T7 endonuclease (T7EI; NEB) assay. 200 ng PCR product was subjected to a denaturing and annealing step. The detailed running parameters are in **Table 2. 8.** Subsequently, 1 µl T7EI enzyme was added into each reaction tube, mixed gently, and incubated at 37°C for 15 min. Then 10 µl was immediately loaded on a 1 % agarose gel, followed by the visualization and band intensity measurement by Image J analysis. The cleaved fraction (f) was calculated a: **f (cut)=a/(a+b).**

In this formula, **a** is the intensity of cleaved products bands, and **b** is the intensity of the undigested PCR product.

Temperature	Time
95°C	10 min
95-85°C	-2°C/s
85°C	1 min
85-75°C	-0.3°C/s
75°C	1 min
75-65°C	-0.3°C/s
65°C	1 min
65-55°C	-0.3°C/s
55°C	1 min
55-45°C	-0.3°C/s
45°C	1 min
45-35°C	-0.3°C/s
35°C	1 min
35-25°C	-0.3°C/s
25°C	1 min
25-4°C	-0.3°C/s
4°C	∞

Table 2. 7 T7EI assay conditions.

TIDE assay

The online tool 'Tracking of insertions and deletions (in/dels) decomposition' (TIDE assay http://tide.nki.nl/) was utilised to quantify the editing efficacy of CRISPR reagents (Brinkman et al., 2014). Following the PCR amplification spanning the target region and sequencing analysis, sequencing trace files of a control and edited samples were input into the online TIDE software. The TIDE software can calculate the cutting efficiency and determine in/del frequencies.

2.3.1.11 Southern blot

The Southern blot was performed according to manufactures instructions and protocol from Edwin Southern (E.M.Southern, 1975). All work involving phenol was carried out in a fume hood, and all steps were carried out at room temperature unless otherwise stated. All homemade solutions in Southern blot are listed **below**:

Solutions	Final Concentration
Cell lysis buffer	10 mM NaCl, 20 mM Tris-HCL, 1 mM EDTA, 0.5% SDS and 100 µg/mL proteinase K (pH=8)
Depurination Solution	2.5 mM HCL
Denaturing solution	0.5 M NaOH, 1.5 m Nacl (pH=13)
Neutralising solution	1M Tris (PH=8), 1.5M NaCl (pH=7.5)
Hybridisation Buffer	0.25 M, Na_2HPO_4, 7% SDS, 1mM EDTA (pH=7.2)
Wash Solution A	2xSSC 2% SDS
Wash Solution B	2 xSSC 1% SDS

Table 2. 8 Southern Blot buffer recipes.

2.3.1.11.1 HMW gDNA extraction for Southern Blot

Prior to the isolation of HMW gDNA, cells were cultured in a T150 flask (Cat#CLS431080-50EA, Eppendorf). Confluent cells were washed, and harvested into a 50 ml falcon tube as described in **Section 2.1.1.1**. Pellets were washed and dissolved in 5 ml lysis buffer. The reaction was mixed and incubated overnight at 55°C. The following day, 5 ml volume phenol (Cat#P1037, Sigma-Aldrich) was added in the lysate, followed by adding phenol-chloroform:isoamyl-alcohol (25:24:1 Saturated with 10 mM Tris, pH 8.0, 1 mM EDTA , Cat#25666, Sigma-Aldrich). The tube was gently inverted, followed by the incubation for 10 min. The reaction was then centrifuged at 12,000 x g for 20 min. The upper aqueous phase containing the nuclei acid was transferred into a fresh tube by a pipette (the top of the tip was cut to be wider), followed by DNA precipitation. 1/10 volume of 3 M sodium acetate (pH=5.5) and 2 volumes of 100% ice-cold ethanol was added into the reaction, and the reaction was mixed well before the incubation overnight at -20 °C. In the following day, the DNA was spooled by a bent, autoclaved and thin glass, followed by a wash with 70% ethanol. The DNA was air dried in a safety hood. Finally, the DNA pellet was dissolved in 100 µl TE buffer and kept at 4°C for long-term storage.

2.3.1.11.2 HMW gDNA restriction digest and visualisation

A 98 µl volume reaction, involving 10 µg HMW DNA, 10 µl of 10x reaction Buffer, 3 µl of the EcoR I restrict enzyme (10 units/µl) and NF water, was incubated overnight at 37°C. On the second day, another 2 µl EcoR I enzyme was added in the reaction, followed by the incubation for 2 h. 20 µl digested sample and 6x loading dye were mixed and loaded on a 0.7% agarose gel electrophoresis. DNA 1 kb ladder (Cat#

15615-024, ThermoFisher Scientific) and Hind III DNA ladder (Cat# SM0102, ThermoFisher Scientific) were loaded as molecular weight markers alongside. The gel was run at 20 V overnight, and followed by the visualisation under a UV-trans-illuminator. A fluorescent ruler was placed alongside the gel and photographed as a measurement of the distance of DNA fragment migrated.

2.3.1.11.3 Transfer

The DNA fragment of interest was larger than 10 kb in size, so the gel was de-purinated prior to the transfer. Briefly, the gel was soaked in denaturing solution for 2 x 15 min, followed by another wash in distilled water. The gel was saturated soaked in the neutralising buffer for 2 x 15 min. The gel was then rinsed in distilled water, followed by the upward capillary transfer.

All components were assembled as shown **below**. Setting a piece of glass plate on a clean glass baking tray and filling the tray with 20 x SSC (CSU, Roslin Institute). Three layers of filter papers (Cat# 88600, Invitrogen) were pre-soaked in 20 x SSC solution, and folded on the glass sheet with their bottom layers overhanging and dipping into 20 x SSC solution. Gently transferring the gel matrix to lay on the top of filter papers, and then placing Saranwrap over the entire glass plate. Cutting a window to fit edges of the gel. A gel-sized piece of Hybond N+ nylon membrane (Cat# 95038, GE Healthcare) was pre-wetted and placed over the gel. The left upper corner was cut and the comb was labelled using a pencil on the membrane. Subsequently, four pieces of pre-wet filter paper the same size as the gel were placed on the top of the membrane. A stack of green towels was placed on top and then covered with a glass plate. A weight was placed on top of stack and left overnight.

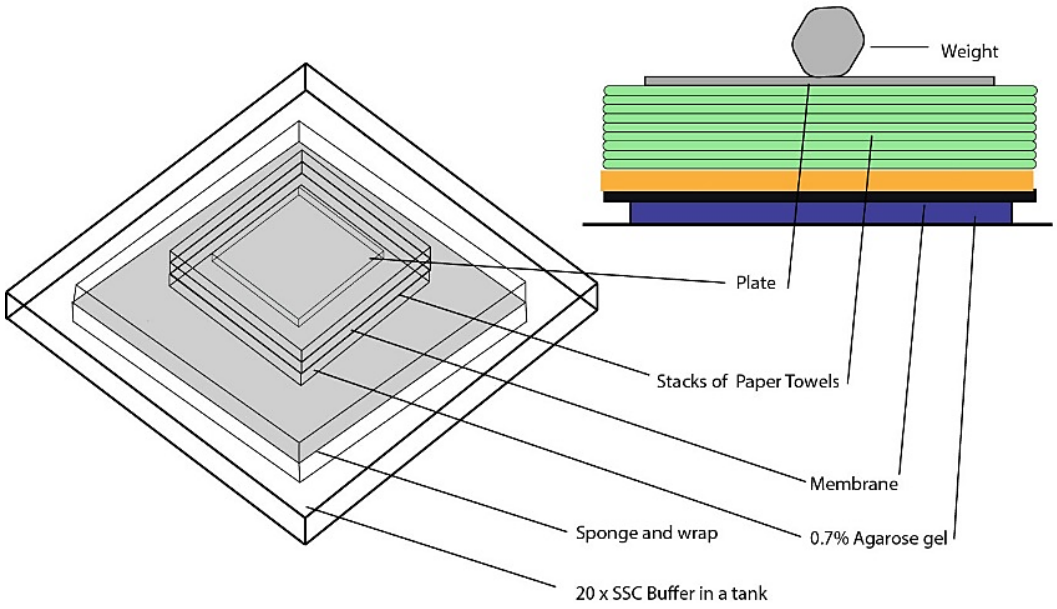


Figure 2. 5 Southern blot transfer assembly.

2.3.1.11.4 Crosslink

On the next day, the blotting stack was dismantled, and crosslinked to the membrane using Stratagene Stratalinker® 2400 UV Crosslinker (Cat#53274-1, Agilent Technologies) for 120 s.

2.3.1.11.5 Probe radiolabelling and hybridisation

A 532 bp DNA fragment was amplified using forward primer (5'-GGGATGGAACCTCAGTCACA-3') and reverse primer (5'-ACCACAGCTACCCACCTTAC-3'). The purified PCR product, as the template, was diluted to 25 ng/µl with MIlliQ water. The final volume was 14 µl. After this, the steps were conducted by Claire Neil (Whitelaw Group). Briefly, the reaction (14 µl) was incubated at 95°C for 5 min and transferred on ice for 2 min. The radiolabelling of the

probe was then carried out by adding 4 µl of High Prime (Cat#. No 11585592001, Sigma) and 2 µl of [α^{32P} dCTP] (Cat# No BLU513H250UC, Perkin Elmer), followed by the incubation for 15 min at 37°C. The unincorporated dNTPs were removed through the spin-column chromatography (Illustra Nick Column, Cat# 17085501, GE Healthcare). The storage buffer from the column was discarded and the column was filled with TE buffer. Another 400 µl TE buffer was added to remove unincorporated [α^{32P} dCTP]. A final 400 µl of TE buffer was poured onto the column, and the flow-through which contained the labelled probe was collected. 30 µl 5 M NaOH was then added to the tube and the reaction was mixed well. The membrane was prepared for hybridisation by incubating in the hybridisation solution over at 65°C for at least 30 min prior to the addition of radioactively labelled probe. The probe was then added in the hybridisation tube and incubated at 65˚C overnight.

2.3.1.11.6 Blot wash, hybridisation and visualisation

Hybridisation solution was removed and the membrane was rinsed in 50 ml wash solution A at 65°C for 20 min, followed by two washes (2 x 20 min) with 50 ml wash solution B. The hybridising membrane was placed on an imaging plate (Fujifilm, BAS-IP MS 2025) and exposed to pre-blanked with GE healthcare Image Eraser 810-UNV for 24 h. Images were then taken using a Typhoon FLA 7000 (https://www.gelifesciences.com/shop/protein-analysis/molecular-imaging-for-proteins/imaging-systems/typhoon-fla-7000-p-00194?current=28955809).

2.3.2 RNA techniques

Please note that all procedures were performed at room temperature unless otherwise stated. Care should be taken throughout the whole procedure to avoid RNA degradation and RNase contamination.

2.3.2.1 RNA extraction, quantification and integrity

Total RNA was extracted from frozen tissues using an RNeasy Mini Kit (Cat# 74104, Qiagen) according to the manufacturer's protocol. RNA samples were quantitated at an absorbance of 260 nm using the NanoDrop spectrophotometer as described (**Section 2.3.1.4**). The RNA integrity (RIN) was calculated using the Agilent 2200 bioanalyzer (Cat#G2963-90002, Agilent Technologies) according to the accompanying protocol. The ratio of 28S:18S ribosomal RNA was estimated.

2.3.2.2 Reverse transcription

cDNA was prepared using the Superscript II cDNA synthesis system (Cat#11904018, Invitrogen) following the manufacturer's protocol. Briefly, 1 µg of total RNA was mixed with 1 µl random primer, 1 µl of 10 mM dNTP. The reaction was made up to 10 µl reaction with NF water, and the mixture was incubated at 65°C for 5 min, before incubation on ice for 1 min. The cDNA synthesis reaction consisting of 4 µl of 25 mM $MgCl_2$, 2 µl of 10 x RT buffer, 2 µl of 0.1 M DTT (1,4-Dithiothreitol, DTT-RO, Sigma-Aldrich), and 1 µl RNaseOUT™ (40 U/ µl) was mixed in a separate tube and added to the 10 µl RNA/primer reaction. The mixture was gently mixed and collected by a brief spin before the incubation at room temperature for 2 min. Subsequently, 1 µl of Superscript™ II RT (50 U/µl) was added. The reaction was

incubated at 42°C for 50 min, and was terminated at 70°C for 5 min. The reactions were then placed on ice. 1 μl RNAse H (2 U/μl) was added to remove RNA, followed by the incubation at 70°C for 20 min. The single-stranded cDNA was used as the template in the quantitative PCR reactions.

2.3.2.3 Quantitative PCR

Quantitative PCR (qPCR) was used to compare the expression levels fo ST6GAL1 mRNA transcripts derived from different pig tissues. To this end, qPCR was performed in a 7500 Fast Real-time PCR System (Cat# 4351107, Applied Bio-system) using the SYBR® Green Master Mix (Cat#4309155, Invitrogen) following the manufacturer's recommendations. The mRNA expression level in each tissue was normalised to housekeeping genes, glyceraldehyde 3-phosphate dehydrogenase (GADPH) and ribosomal protein L4 (RPL4), which have been validated to be stable expressed in pig tissues (Nygard et al. 2007). The specific primers of ST6GAL1 gene and housekeeping genes used are detailed:

Primer	Sequence (5' - 3')
ST6GAL1_Forward	CTGCATCCTGATCAGCCCTT
ST6GAL1_Reverse	CTGGTCACACAGCGTCATCA
GADPH_Forward	ACACTCACTCTTCTACCTTTG
GADPH_Reverse	CAAATTCATTGTCGTACCAG
RPL4_Forward	CAAGAGTAACTACAACCTTC
RPL4_Reverse	GAACTCTACGATGAATCTTC

Table 2. 9 Primers for qPCR.

In each cycle, the fluorescence intensity was determined by the binding of SYBR green dye and dsDNA, allowing the concentrations of samples quantified. Each reaction was in triplicate and the specificity of the amplification was determined by the analysis of melting curves. The final concentration of each component and the following thermal cycling conditions are described as **below**:

Component	Volume (uL)	Final Concentration	
2 x SYBR® green Master Mix	10	1x	
Forward primers	5	200 nM	
Reverse primers	5	200 nM	
cDNA	Variable	1 - 100 ng	
Distilled water	Up to 20 µl	-	
Procedure and parameters	Temperature	Time	
Activation	95°C	10 min	
Denature	95°C	10 s	40cycles
Anneal/Extend	60°C	1 min	

Table 2. 10 Conditions of qPCR using SYBR® Green MasterMix.

2.3.3 Protein techniques

All procedures were performed at room temperature unless otherwise stated.

2.3.3.1 Western Blot

All procedures were performed according to the protocol from Prof. Paul Digard group (Digard et al., 2017). All antibodies (Table 2.) and solutions used are listed below:

Antibodies	Supplier/Sources
Primary antibody: Rabbit polyclonal Anti-myelin basic protein (MBP)-NP (2915) (1:1000 dilution)	(Turnbull, 2017)
Primary antibody: Rat monoclonal Anti-α-Tubulin (1:1000 dilution)	Serotec (MCA77G), clone YL1/2
Secondary antibody: IRDye® 800CW Donkey Anti-Rabbit IgG (H + L) (0.5 mg, 1:10,000 dilution)	P/N 926-32213 LI-COR Corporate Offices
Secondary antibody: IRDye® 680LT Donkey Anti-Mouse IgG (H + L) (0.5mg, 1:10,000 dilution)	P/N 926-68022 LI-COR Corporate Offices

Table 2. 11 List of antibodies.

Buffer	Components and Supplier
Laemmli's solution	20% (v/v) Glycerol 2% (w/v) SDS 100 mM DTT 24 mM Tris 0.016% (v/v) brompophenol blue 0.016% (v/v) xylene cyanol solution
4 x Resolving buffer (Protogel)	0.375 M Tris-HCl 0.1% SDS pH 8.8 (Cat#EC-892, ProtoGel)
Resolving polyacrylamide gel	30% acrylamide:bisacrylamide (37.5:1) 4x resolving buffer 10% (w/v) APS (Cat#09913, Sigma-Aldrich) TEMED (Cat# T7024, Sigma-Aldrich)
4 x stacking buffer (Protogel)	0.125 M Tris-HCl 0.1% SDpH 6.8 (cat#EC-893, ProtoGel)
Stacking polyacrylamide gel	30 % acrylamide:bisacrylamide (37.5:1) 4 x stacking buffer 10% (w/v) APS TEMED
SDS-PAGE running buffer (Bio-rad)	0.25 M Tris 1.92 M Glycine 1% w/v SDS
1x transfer buffer	0.025 M Tris 1.92 M Glycine 0.1% (w/v) SDS 25% % (v/v) Ethanol
Blocking solution	10% (w/v) Semi-Milk Powder 0.1% % (v/v) Tween 20/PBS
Wash buffer	0.1% (v/v) Tween 20 1 x PBS

Table 2. 12 Buffers and solutions used in western blot.

Confluent cells cultured (4 x 10^4 cells) in 24-well plates were lysed with 200 µl Laemmli's solution, followed by being re-solubilised at 95°C for 5 min, and centrifuged at 12,000 x g for 5 min. Samples were kept at -80°C or loaded on the sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (Cat#4561086, Bio-Rad). Mini-PROTEAN tetra system (Cat#1658035, Bio-Rad) was used to for SDS-PAGE apparatus. After loading up to 15 µl in each lane of the SDS-PAGE, the tank was filled with 1x SDS running buffer. The gel was run at 80 V for 10 min, and then switched to a constant voltage of 120 V for approximately 1 h. The colour pre-stained protein standard with a broad range (Cat#P7712S, NEB) was used as a marker. Protein was then transferred from the SDS-PAGE to a nitrocellulose membrane using trans-Blot Turbo™ Transfer system (Cat#1704150, BioRad) according to the guideline of manufacture. Then the cassette was disassembled and the membrane was blocked in 5%(w/v) skimmed milk (Marvel) in PBST/0.1%(v/v) Tween 20 for 30 min.

All primary antibodies were diluted in 5% (w/v) skimmed milk in PBST and membrane incubations were conducted on an orbital shaker for 1 h at room temperature or overnight at 4°C. Each blot was probed with a mixture of a primary antibody and a loading control at proper concentrations, followed by throughout three washes in PBST. Then the blot was incubated in a secondary fluorescent tagged Li-COR antibody at room temperature for 60 min in the dark, followed by three washes. The visualisation of the blot was performed using Li-COR Odyssey imaging system (Lot#410, Li-COR Bioscience), according to the manufacture's protocol. The band intensity was quantified with Image J.

2.3.3.2 Lectin staining

All procedures were performed at room temperature unless otherwise stated. Fluorescein labelled lectins, *Sambuccus nigra* agglutinin (SNA) lectin, *Maackia amurensis* lectin (MAL II) and *Canavalia ensiformis* (ConA) lectin (Vector Laboratories) were used to detect α 2,6-linked SA, α 2,3-linked SA, and α-mannose activity on cell surfaces, respectively.

For the confocal imaging, cells were pre-seeded on the coverslips in 24-well plates. Cells at a confluence of ~50% in a 24-well plate were washed in PBS and fixed with 250 µl 4% formaldehyde in PBS for 20 min. After this, cells were washed in 1% FBS/PBS (3 x 5 min), and then permeabilized with 250 µl of 0.2% Triton X-100/PBS. After washes (3 x 5 min) with 1%FBS/PBS, cells were blocked with 1% FBS/PBS for 1 h, before the incubation with 50 µl of SNA, MAL II or ConA lectin solution (10 µg/ml) for 1h or overnight at 4°C. Cells were then washed with 1%FBS/PBS, before the incubation with avidin-FITC conjugate (10 µg/ml) and Hoechst DNA dye (0.5 µg/ml, Cat#62249, ThermoFisher Scientific Inc.) for 40 min in the dark. Final washes (3 x 5 min) were developed with 1x PBS before mounting and visualisation. The confocal laser scanning microscope (LSCM) coupled with Z-stack were used to capture images at focal distances according to the manufacture's protocol. In this study, up to Z-stacking images at 80 µm vertically were captured and aligned under 63x objective lenses. The wavelength of FITC-A channel used was 488 nm.

For quantitative analysis of surface SA expressions, confluent cells were seeded in 24-well plates and harvested as described in **Section 2.1.1.1**. The procedure of lectin staining was performed essentially same as shown above. More than 2,000 stained cells were transferred to FACs tubes for flow cytometry analysis with a laser

violet 405 nm, which was according to the same instructions as previously described (**Section 2.2.2.4.3**).

2.4 Statistical analysis

Experiments were analysed with GraphPad Prism v6 and Microsoft Excel software. Data was presented as mean±SEM. A one-way analysis of variance (ANOVA) was used to determine the statistically significance, and p-values <0.05 was regarded as statistically significant.

3 Identification of 5' untranslated regions of pig ST6GAL1 mRNA transcripts

3.1 Introduction

β-Galactoside α 2,6-Sialyltransferase (ST6Gal1) is known to be an important sialyltransferase (ST) responsible for the addition of terminal sialic acid (SA) residues to Gal 1,4 GlcNAc disaccharide on N-glycans or O-glycans with α 2,6-linkage (Kitazume, 2014), termed as 'α 2,6-sialylation'. Aberrant α 2,6-sialylation of glycoproteins has been reported to be associated with multiple physiological and pathological aspects, such as cell apoptosis (Swindall and Bellis, 2011), cell adhesion (Suzuki and Abe, 2014), and inflammation (Nasirikenari et al., 2010). Moreover, the upregulation of α 2,6-sialylation on the cell surface is characteristic of malignant carcinogenesis, such as in gastric cancer (Gretschel et al., 1998), liver cancer (Dall'Olio et al., 1999), breast cancer (Minami et al., 2013) and colorectal cancer (Zhang et al., 2017b). Importantly, in mammalian immune system, ST6GAL1 is highly expressed in B cells (Jones et al., 2016) and α 2,6-linked SA can act as a ligand for CD22 in B-cell receptor activation as evidenced in ST6GAL1 homozygous deficient mice (Collins et al., 2006). It is also implicated in host-pathogen interactions for several diseases (Ogata et al., 2009), such as influenza A virus (IAV), which is the interest of this project.

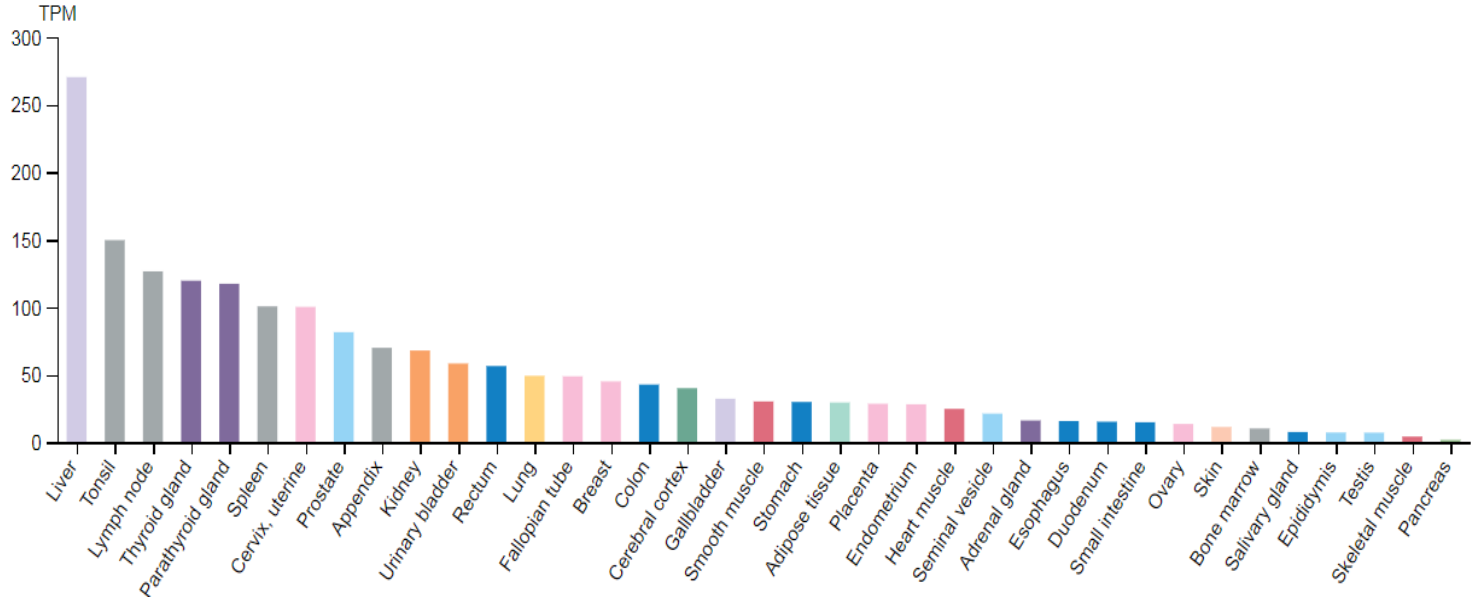


Figure 3. 1 Human ST6GAL1 mRNA expression levels in different tissues. The HPA data was obtaned from https://www.proteinatlas.org/ENSG00000073849-ST6GAL1/tissue . TPM stands for transcripts per million.

ST6Gal1 is encoded by the ST6GAL1 gene, which has been extensively studied in humans (Svensson, 1990), mice (Hamamoto et al., 1993), cattle (Mercier et al., 1999), rat (Weinstein et al., 1987), and chicken (Kurosawa et al., 1994). Taking human ST6GAL1 as a reference in the Human Protein Atlas (Uhlen et al., 2015), it is expressed differently from tissues to tissues (**Figure 3. 1**). This RNA-seq data is corresponding with the previous study, in which mammalian ST6GAL1 mRNA was detected in liver with the high levels (Petit et al., 2010).

Regulation of ST6GAL1 expression is mainly at the transcriptional level. The different expression of ST6GAL1 across tissues is regulated by multiple promoters (Wang X 1993), generating several transcripts encoding the identical polypeptide (T P O'Hanlon, 1989). The multiple transcripts of ST6GAL1 in humans, mice, rat and cattle differ in their 5' heterogeneous untranslated region (UTR) sequences. Totally, there are 22 transcript variants of ST6GAL1 mRNA reported in the human genome assembly (GRCh38.p12), six in the mice genome assembly (GRCm38.p4), two in the bovine genome assembly (ARS-UCD1.2), three transcripts in the rat genome assembly (Rnor_6.0), and two in chicken genome assembly (GRCg6a). However, compared with them, there was only one transcript variant in the pig assembly (Sscrofa 10.2), and a lack of 5' UTR and detailed annotations of the transcriptional profiling.

ST6GAL1 mRNA has three transcript variants (forms 1, 2 and 3) identified (Maksimovic et al., 2011), governed by at least three physically and independent promoter regions characterized (Lo and Lau, 1999), P1, P2 and P3. P1 sequence (NC_000003.12) generates mRNA form 1, which contains 5' UT exons (UTE) H and I (Mercier et al., 1999). P1 activity has been characterised in liver of adult animals (YP Hu, 1997), involved in acute phase response (Kaplan HA 1983). P1 region is also

identified in colon and intestinal cells (Charles W. Hutton, 1987). The loss-of-function study showed that the ablation of ST6GAL1-P1 region resulted in strikingly depressed levels of ST6Gal1 levels in circulation (Jones et al., 2010). Moreover, ST6GAL1-P1 deficient mice exhibit a normal B cell response, but a great burden of the pathogen *Salmonella typhimurium* in liver and spleen (Appenheimer et al., 2003b).

P2 sequence (NC_000003.12) promotes the expression of ST6GAL1 mRNA form 2, containing 5' UTE X, H and I (Lo and Lau, 1996). P2 region has been reported to contain a specific locus involving *cis-acting* regulatory elements responsible for the promoter activity in B-lineage cells (Lo and Lau, 1996), and thus it is important in B-lymphocytes maturation (Appenheimer et al., 2003b). In particular, E(X) has been reported to be associated with with repressed expressions in a B-null precursor cell line and human liver cancer (HepG2) cell lines (Lo and Lau, 1999).

P3 sequence contributes to the basal expression of mRNA form 3, which contains UTEs Y, Z, H and I (Dall'Olio, 2000). P3 region contains CpG islands, as short interspersed DNA sequences that are guanine-cytosine (CpG)-rich, is typically close to the transcription start site (TSS) of ST6GAL1(Bird, 2011). However, P3 region is associated with a non-tissue-specific expressed pattern of ST6GAL1 (Svensson, 1990), and responsible for the expression in other cell types (Appenheimer et al., 2003a) in which P1 and P3 are inactive.

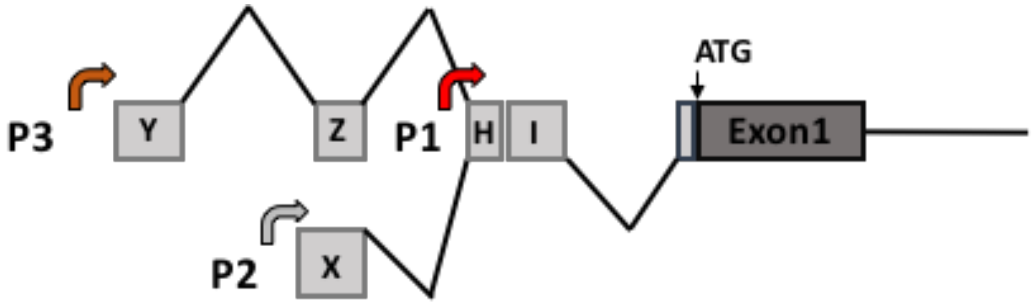


Figure 3. 2 Diagrammatically representations of ST6GAL1 mRNA transcripts in human, mice, rat and cattle. Associated promoter regions (P) are denoted by arrows, P1 is orange, P2 is grey and P1 is red. The data were summarized from the transcriptional profiling of ST6GAL1 in human, bovine and rat (L Meng et al., 2013). Translated exons are black boxes and untranslated ones in grey.

It has been reported that the useage of diverse promoters in ST6GAL1 gene are evolutionarily retained in human, mice, cattle and rat ST6GAL1 (Appenheimer et al., 2003b). Morover, cross-species comparisons of conserved non-coding sequences enable one to identify functional regulatory elements (Dubchak and Frazer, 2003). Therefore, the study of transcriptional profiling of pig ST6GAL1 mRNA enables us to gain a better understanding associated transcriptional mechanism expected from 5' UTR in different pig tissues and organs. To this end, we perform reverse genetics techniques to investigate 5' alternative splicing variants of pig ST6GAL1 mRNA. We then grouped those transcripts into various mRNA families, which differ by their 5' divergent regions. The study provides resources to ST6GAL1 gene functions involved in a range of biological systems, especially its implications in infectious pathogen virulence (Shi et al. 2017).

3.2 Results

3.2.1 Revision of exon 1

In the pig genome assembly Sscrofa 10.2, intron 1 in the coding region of ST6GAL1 mRNA contained a region of roughly 100 bp indicated by 'N's (**Figure 3. 3**). We performed PCR using primer F3 and R3, using an optimized annealing temperature 68°C, followed by sequencing. Sanger sequencing revealed 'AAG' as the unknown gap between known exon 1 (E1) and exon 2 (E2). As this is too short to be a genuine intron (Jo & Choi 2015), these two exons were combined into a new exon re-defined as 'exon 1'. We now have a complete and correct sequence of the first exon for ST6GAL1.

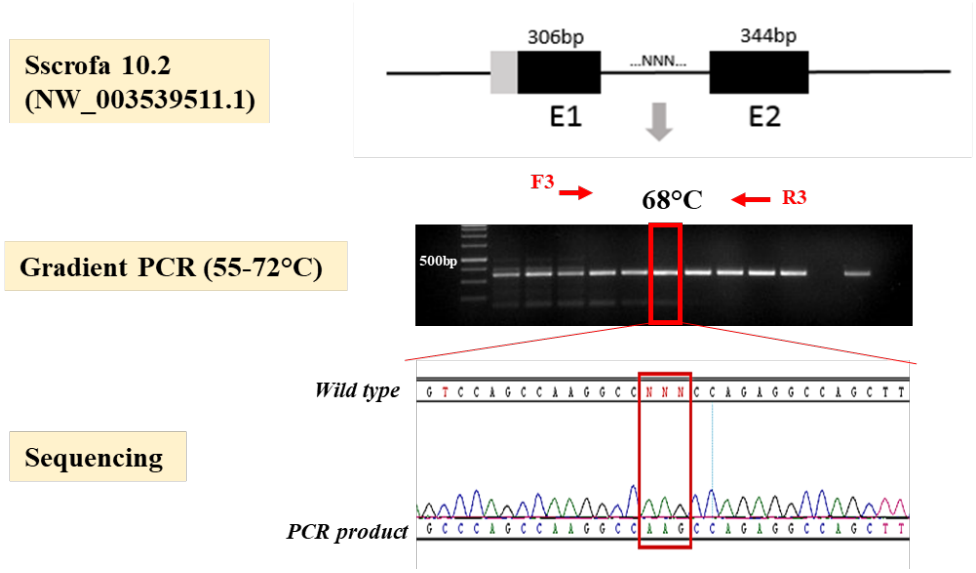


Figure3. 3 Reconstruction of coding region in exon 1 sequences. The agarose gel exhibits the products, which were amplified by PCR using primers F3 and R3 surrounding the unknown region between exons 1 and 2 (E1 and E2, black boxes). The unknown sequence between E1 and E2 was 'AAG'.

3.2.2 Assessment of RNA integrity

RNA, extracted from nine pig tissues-lung, trachea, bronchus, liver, kidney, spleen, stomach, heart and small intestine as described **(Section 2.3.2.1)**, was assessed of RNA integrity (RIN) (Wilkes et al., 2010). RIN is commonly used to assess RNA quality. RNA quality is considered to be sufficient if the RIN value is greater or equal to 7. Low quality of RNA may compromise the results of downstream experiments (Schroeder et al., 2006). The metric scoring of 10 RIN means a lower level of degradation and >=7 guaranteed for transcriptomic analysis. The ratio of 28S and 18S ribosomal components was also measured. The pseudo electropherograms in which the total RNA bands at 28 and 18S ribosomal RNA (rRNA) bands are clearly distinguishable in the electrophoretogram, indicating that the extracted RNA was of good quality.

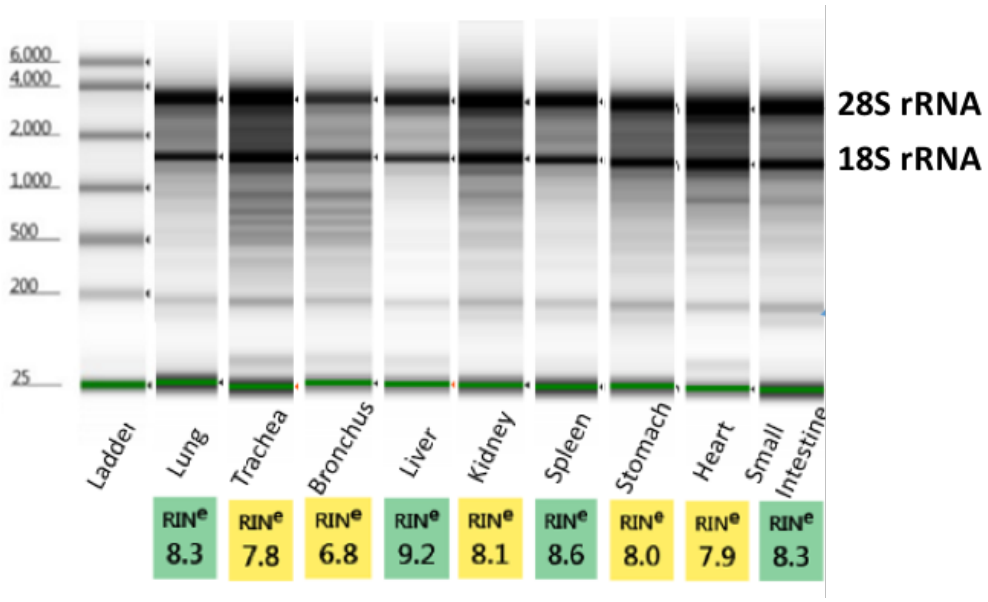


Figure 3. 4 Assessment of RNA integrity. Results of the RNA assessment of nine tissues using the Agilent Bioanalyser 2100. The 28S and 18S RNA subunits are visible at ~4000 bp and ~2000 bp.

3.2.3 ST6GAL1 mRNA expression in various tissues

ST6GAL1 mRNA expression in various tissues was determined by qPCR (**Section 2.3.2.3**). As can be seen, amongst all the ST6GAL1 mRNA transcripts detected, ST6GAL1 was most abundant in the liver tissue, which interestingly was also observed in the human ST6GAL1 levels (**Figure 3. 1**). Therefore, we hypothesis that the tissue-specific expression pattern of pig ST6GAL1 mRNAs is associated with 5' splice variants and differential utilization of regulatory elements, which is described in the later sections. This is the first study focused on tissue-specific expressions of pig ST6GAL1 mRNA.

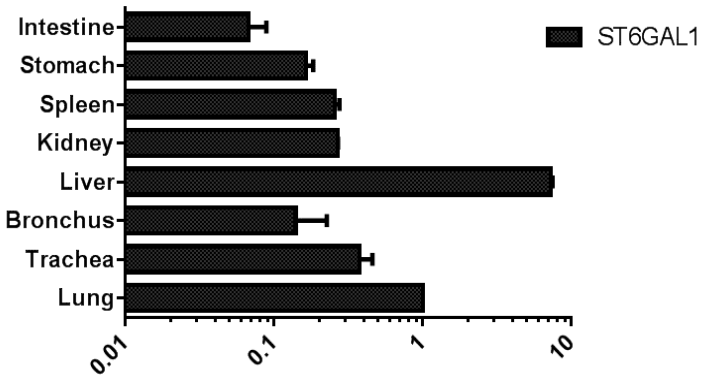


Figure 3. 5 Relative expressions of ST6GAL1 mRNA in different pig tissues. qPCR results show ST6GAL1 expression levels from various pig tissues. Data is normalised to housekeeping genes, GADPH and RPL4. The value of expression in lung is set as '1'. Data is expressed as mean with standard error, n=3. x-axis: relative expression levels.

3.2.4 Identification of 5' UTR of ST6GAL1

5' RACE was carried out to obtain 5' UTR sequences of ST6GAL1 mRNA in various pig tissues. Prior to the experiment, the positive control provided with the kit was used to optimise the 5' RACE system. The template used in this experiment was an *in vitro* transcribed RNA of the chloramphenicol acetyltransferase gene (**Appendix 1).**

The 5' RACE of ST6GAL1 transcripts was performed as described in **Section 2.3.1.3.2.** The nested PCR visualisation on the 2% agarose gel is shown in **Figure 3. 6a.** This gel image was a representative one in three independent repeats. The supplementary results are in **Appendix 2**. Each tissue produced at least one cDNA amplicon, whose size ranged from 200 bp to 600 bp. The PCR amplicons with robust enrichment were marked with yellow stars. Each individual PCR product was numbered in a descending order according to its size.

All these PCR products obtained were extracted, and sub-cloned into the pCR™4-TOPO® vector, followed by sequencing from M13 reverse primer. Each result showed the liner sequencing inserted into the vector. Taking spleen 1 transcript as an example, the sequencing trace file is shown in **Figure 3. 6b**. The sequence was treated as a distinct entity amplified between the 5' anchor primer and nested primer (NP). More specifically, it was the sequence between the 5' transcription start site (TSS) and the known sequence (~250 bp downstream of ATG start codon). The sequence obtained from each sample was then aligned to the pig assembly 10.2, showing that they were made up of different untranslated exons (UTE). We thus named them from E(-1) to E(-9) (3' to 5'). Sequence of all nine UTEs was detailed in

Appendix 3. However, the sequence of lung 1 (**Appendix 2**) and heart 1 (**Figure 3. 6-a**) transcripts was failed.

To better illustrate the results and to more clearly characterise the pattern of 5' spliced variants, transcripts were grouped into different mRNA families according to the 5' divergent sequence. Ten mRNA families were isolated from 24 ST6GAL1 transcripts. The schematic representation of these variants is shown in **Figure 3. 7** (**upper panel**). The most abundantly expressed transcripts in each tissue marked with asterisks.

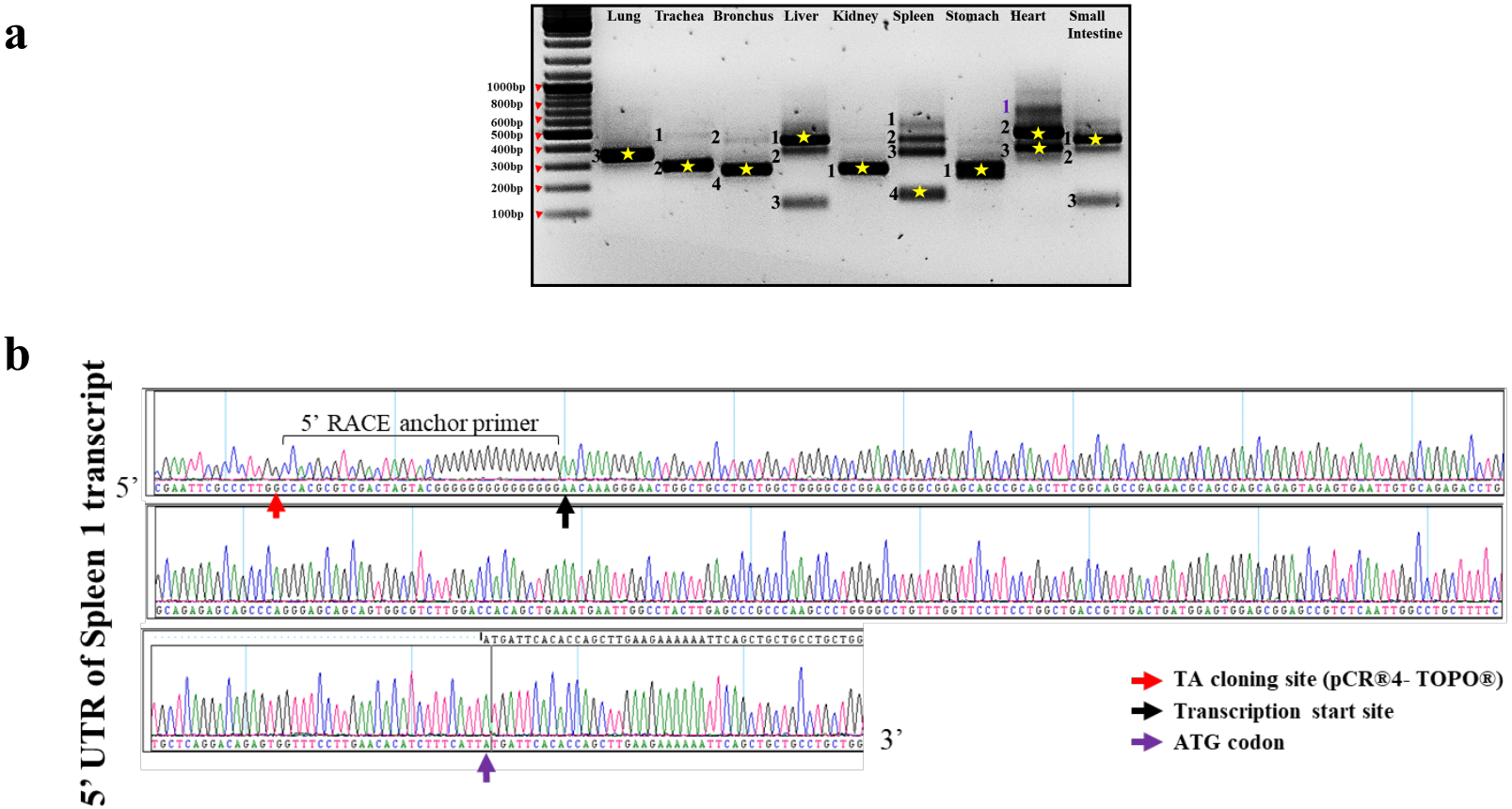


Figure 3. 6 5' RACE results and sequencing. In each lane, products were numbered according to the sizes in descending order. The gel was the most representative result, and other numbered products accordingly were in other repeats. Yellow asterisks indicate the most abundant transcripts from each tissue.

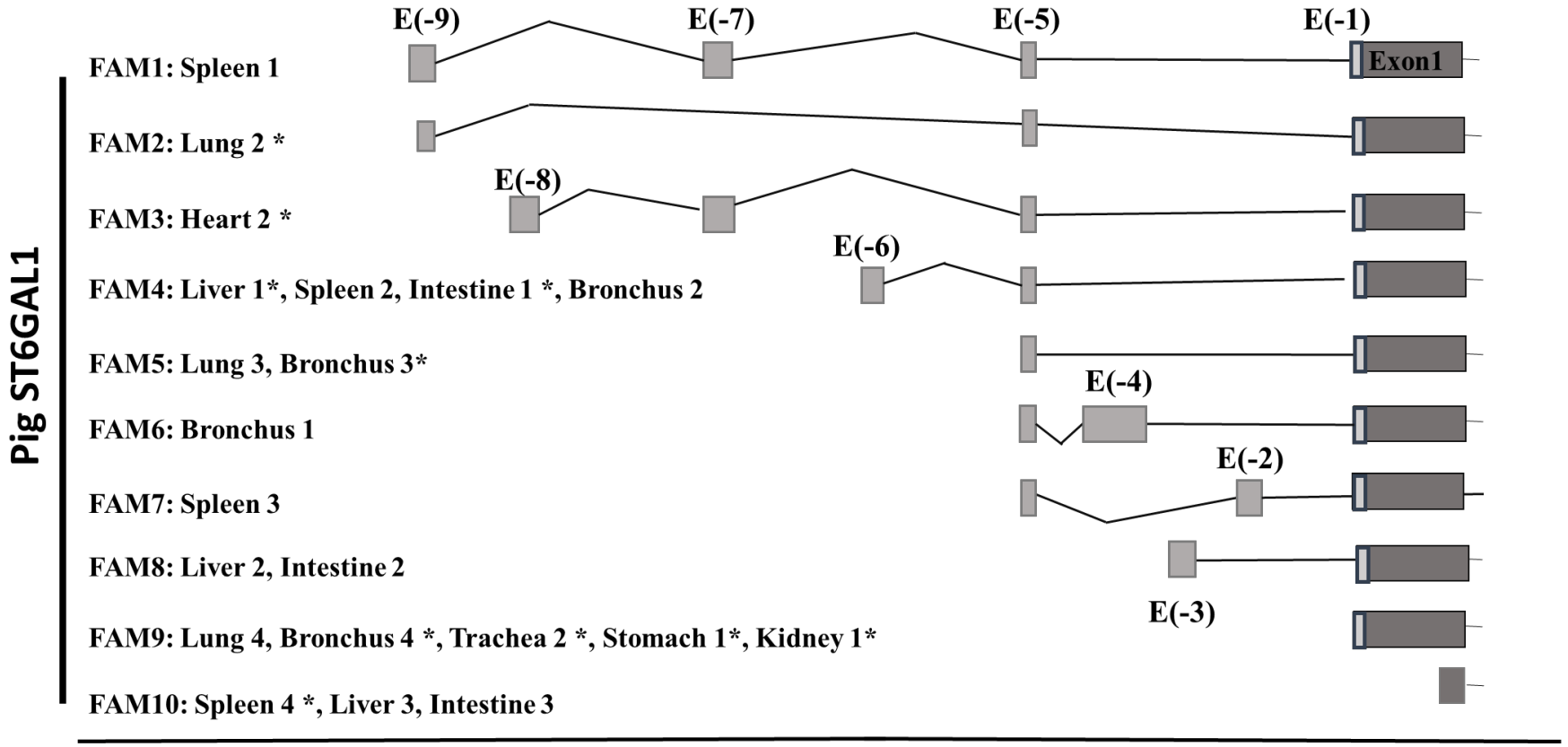


Figure 3. 7 Schematic representation of pig ST6GAL1 transcripts in nine tissues. Translated exons are shown in dark grey and UTEs are in light grey. Within each family, the asterisk indicates the most abundantly expressed transcript for that tissue. Negative numbers are used to record UTEs.

3.2.5 Sequence alignment

The experimental findings of the transcriptional profile of ST6GAL1 were further validated by the computational prediction and now form a part of new pig assembly Sscrofa 11.1. The comparison of two datasets shown below suggests that 5' RACE performed in this study is an efficient method to identify transcript variants.

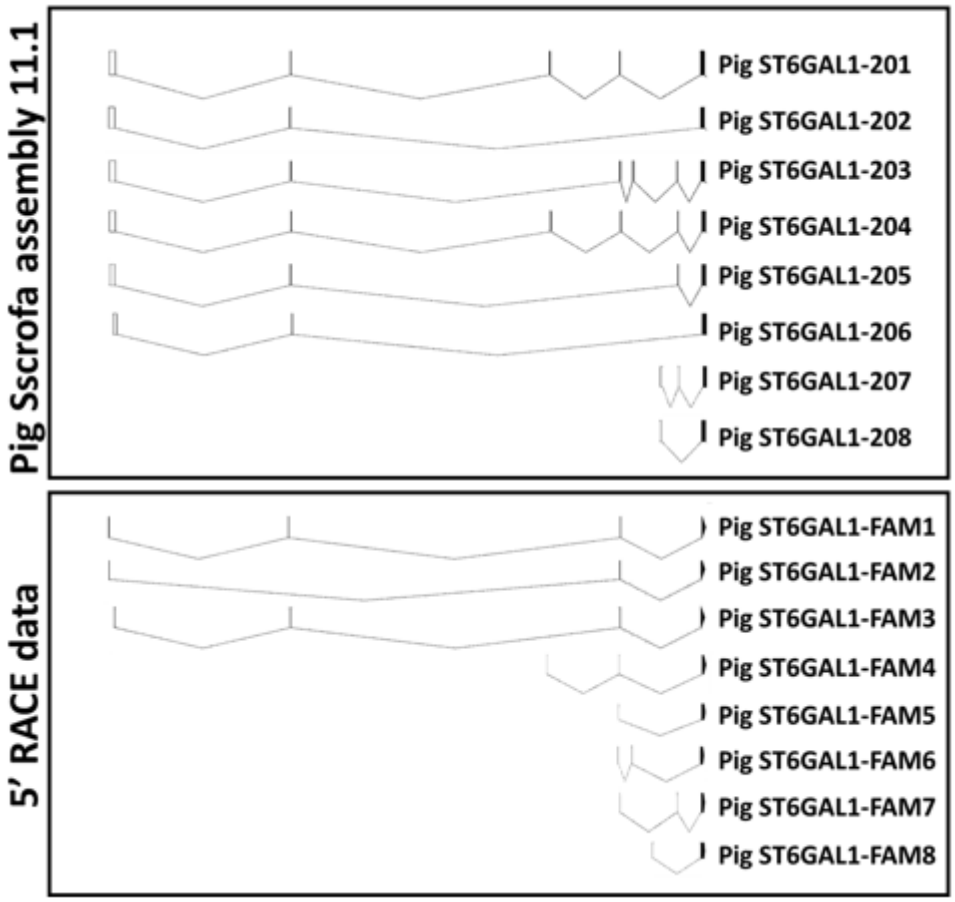


Figure 3. 8 Pig assembly 11.1 versus 5' RACE data. Upper image shows the schematic map of pig ST6GAL1 mRNA transcripts from pig genome assembly 11.1 that was released after we got RACE data.

3.3 Discussion

The ancestral profile of the ST6GAL1 gene has been reported to be conserved in human, cattle, mouse and rat (Mercier et al., 1999). Moreover, the tissue-/cell-specific expression manner of ST6GAL1 is determined by distinct usages of 5' UT exons and multiple promoter regions (Dalziel et al., 2001). Following these early studies, our work focused on the identification of 5' UTEs of pig ST6GAL1 transcripts, and obtained 24 spliced variants, which were grouped into 10 forms. These variants differ by their diverse 5' TSSs.

The transcriptional profiling of human ST6GAL1 has been previously obtained by homologous evolutionary alignment among species orthologues (Maksimovic et al., 2010). Based on this, we aligned the DNA sequence of pig ST6GAL1 E(-5), E(-7) and E(-9) sequence with human ST6GAL1 E(H), E(Z) and E(Y), respectively (see **Appendix 4**). Moreover, we aligned 500 bp upstream of pig ST6GAL1 E(-5) with human P1 region (see **Appendix 5**). We observed similar regulatory motifs, suggesting that E(-5) expression might be driven by a functional promoter. This hypothesis was supported by a bioinformatics analysis performed by Dr. Sylvia Beka, a visitor in the Whitelaw group. Dr.Beka used several software (described in **Section 2.3.1.9**) to predict promoters associated with each mRNA family (see **Appendix 6**). The study helped us characterise alternative splicing patterns of ST6GAL1. It is known that alternative splicing is closely associated with the regulation of protein functions, and thus is an important mechanism in physiological processes, such as immune response, tumorigenesis, and cell apoptosis (Möröy and Heyd, 2007). Regarding the splicing pattern of pig ST6GAL1 transcripts, we found that the expression pattern was similar

to that of human ST6GAL1, suggesting that pig ST6GAL1 might be associated with a wide range of biological functions.

We now discuss several technical limitations of our experiments. The first limitation is in obtaining full-length cDNA copies from mRNA. The reason is as follows. The homopolymeric tails added on the premature and mature cDNAs are similarly efficient. Therefore, cDNAs made up of large amounts of premature cDNAs cannot be efficient amplified. Also, high GC content at 5'ends could inhibit reverse transcription. Another one is the unstable performance of PCR (Chen et al. 2016; Liu et al. 2002). This can be improved by the optimisation of PCR conditions, as well as increasing amplification cycles.

Yet another limitation is the inefficiency of the DNA sequencing technique to obtain the long DNA sequence due to inefficient primer binding, or DNA degradation, or sample contamination. For example, with lung 1 and trachea 1, we obtained PCR product but failed to obtain their sequencing data. A possible improvement is to apply a 'walking primer' strategy in which the successive PCR primers are used to obtain the 5' or 3' amplicon ends. Another limitation is to distinguish the splicing donor and acceptor sites of UTEs because the sequencing results we obtained were the combinations of different UTEs.

To characterise the tissue-specific expression pattern of ST6GAL transcripts, there are alternative methods, such as Northern blot and high-throughput RNA-seq. Northern blot can exclusively detect tissue-specific mRNA transcripts in RNA mixed pools. RNA-seq is a versatile and unbiased technique performed by next-generation sequencing (NGS). RNA-seq data allows a digital identification of transcript expression profiles derived from diverse tissues (Castle 2014). Alternatively, another

method, serial analysis of gene expression (SAGE), could be applied. In SAGE, a library with short sequence tags can be constructed. SAGE allows transcripts to be assessed according to on the tag frequency (Brufsky 2003). Among all possible techniques, 5' RACE was chosen for the following reasons. Compared with other options, it is cheaper, more accessible and less time-consuming. 5' RACE allows us to explore and isolate unusual transcripts of ST6GAL1 and characterise the 5' cDNA ends in tissues of interest. Unlimited numbers of clones can be generated using 5' RACE. The spliced variants can be visualised, isolated and cloned separately.

To the best of our knowledge, this study is the first one to obtain multiple pig ST6GAL1 transcripts across nine pig tissues. Moreover, the tissue-expressed pattern is revealed, which has critical homologies with that in human ST6GAL1. What we discovered is consistent with those in the newly released pig assembly 11.1. The data provide resources to decipher the tissue-associated expression of ST6GAL1 in diverse tissues and cells, which is more complicated than previously discovered.

Important regulatory elements are commonly located around 5' TSS, and upstream 200 bp proximal to the TSS is typically referred to as core promoter (Lee et al. 2012). Therefore, we will manipulate 5'TSSs in airway-abundant transcripts, which will be detailed in **Chapter 4**.

3.4 Conclusion

We obtained 5'UTR sequences in nine pig tissues. Then we grouped them into ten mRNA families depending on their unique or similar 5'TSS. The transcriptional profile of pig ST6GAL1 gene was thus constructed.

4 ST6GAL1 editing in pig trachea cells

4.1 General introduction

4.1.1 DNA damage and repair

During the life cycle of an organism, DNA in the cells can be damaged due to various exogenous or endogenous factors (Jin et al., 2016). DNA damage, in turn, impacts a wide range of biological processes encompassing mutagenesis, carcinogenesis, normal progression of cell replication and ultimately lifespan (Lieber, 2010).

DNA damage occurs when its structure changes as a result of oxidative damage or DNA hydrolysis (Gates, 2009), hydrolytic deamination (Dabney et al., 2013), O6-methyl-guanines(Srivenugopal and Ali-Osman, 2002), exposure to high-energy radiation or DNA replication fork collapse (Rodgers and McVey, 2016). DNA damage can result in DNA lesions, intra-/inter-strand crosslinks, and single-/double-stranded DNA breaks (Taylor et al., 2015). Among these varieties, double-stranded DNA break (DSB) is a relatively common event. The frequency of DSB-induced events in rats, for example, is estimated to be around $7.4x10^{-5}$/cell/day (Scott and Pandita, 2006). Therefore, DSB is arguably the most severe hazard to genome integrity, potentially causing local mutations, chromosome rearrangements or abnormal numbers of chromosomes in cells (Cheng et al., 2012).

Cells deal with this damage by inducing various repair mechanisms, termed as ‘DSB repair (DSBR)’. When DNA is repaired, the broken DNA ends are bridged together between the gaps, and stochastic insertions or deletions (indels) are sometimes introduced into the genome at the break site. Two principal pathways have been well

studied for DNA DSB repair (DSBR) in eukaryote cells, non-homologous end-joining (NHEJ) and homologous recombination (HR). The critical bifurcations of them are the fidelity of the repair, the requirement for a template and the initial processing of DNA ends (Symington, 2014).

4.1.1.1 Non-homologous ending joining (NHEJ)

NHEJ is an error-prone mechanism that does not require extensive sequence homology in order to recombine the DNA ends. In most eukaryotic cells, the predominant pathway of DSBR is canonical NHEJ (C-NHEJ) (Chiruvella et al., 2013). It can occur throughout the cell cycle. In brief, when a DSB is formed, the two diffusing ending points of the dissected DNA overhangs are stabilised through binding by the ring-shaped Ku heterodimer, preventing end resection and recruiting other factors such as DNA-dependent protein kinase catalytic subunit (DNA-PKcs). This results in the tethering of DNA ends into a synaptic complex (Agrawal and Schatz, 1997), followed by the recruitments of other essential factors, such as the complex of X-ray cross-complementing protein 4 (XRCC4) (Yano et al., 2008) and DNA ligase IV (Costantini et al., 2007) to the site of the DSB. The conjunction ends are then cleaved, followed by the final ligation at the gap of DSB. DNA alterations can occur during these processing steps, resulting in the introduction of small indels at the sites of breaks.

DSBs can be also repaired through one or more alternative NHEJ (alt-NHEJ) mechanisms (Deriano and Roth, 2013), such as single-strand annealing (SSA) and microhomology-mediated end joining (MMEJ). In SSA pathway, DSB ends are bridged by homologous and complementary repeats (more than 25 nt), causing a intra-

chromosomal deletions and/or DNA rearrangements in mammalian cells (Sallmyr and Tomkinson, 2018). MMEJ initiates when DNA breaks are unwound, resulting in exposure of short, single-stranded micro-homologies. These are annealed, the 3' overhangs are cleaved, and the DNA gaps are filled resulting in a variable level of genetic information loss (Sekelsky, 2017). Compared with C-NHEJ, MMEJ is Ku motif independent and increased during the S phase of the cell cycle.

4.1.1.2 Homologous Recombination (HR) and Homology-Directed Repair (HDR)

Another critical pathway for repairing DSB is HR, which occurs in G2 and late S phases of the cell cycle (Langerak and Russell, 2011). During HR, cells use the intact complementary strand DNA or sister chromatid as a template to repair the break without introducing errors. DNA ends are processed and 3'-ends bound by Rad51, which is a DNA-dependent ATPase and strand-exchange protein. Strand invasion occurs when this nucleoprotein filament invades the intact DNA template (ssDNA) (Liu et al., 2011), followed by the establishment of a connection between the invading DNA strand and intact DNA template. A Rad51-dsDNA hetero-duplex (D-loop formation) is generated. In the final step, the invading 3'-ends acts as a primer and Rad51 is released from the dsDNA, followed by DNA strand extension. Therefore, the repair is completed.

HDR has been utilised to introduce exogenous DNA templates at the site of DSB. It occurs during S/G2 phases and M phase of cell cycle, during which sister chromatids are naturally positioned to be the templates (Zaboikin et al., 2017). Donor DNA can be either single-stranded oligodeoxynucleotide (ssODN) (from 40 bp up to several Kbps) or double-stranded DNA (dsDNA) containing a modified sequence

flanked by homologous arms from the targeted locus. When the DSB event occurs, a globular protein, the meiotic recombination 11 (MRE11) can bind to the ends of DNA, where ataxia-telangiectasia mutated (ATM) signalling is activated (Langerak et al., 2011); thus ATM regulates DNA resection, with the generation of small deletions at the 5' terminal DNA.

HDR pathway occurs in a very low frequency (<5%) (Cong et al. 2013). To enhance HDR integration into the target locus, several strategies have been investigated, such as the selection of specific cell and tissue types, cell cycle manipulation or careful selection of target locus (Miyaoka et al. 2016). For example, NHEJ is naturally suppressed during S/G2 phase, resulting in a relative increase in the probability of an HDR event. For this reason, cells can be treated with DNA polymerase inhibitors so that cells are arrested in S/G2 phase due to the inhibition of microtubule multimerisation and integration (Hu et al., 2009). It has been reported that the inhibition of DNA ligase IV results in a significant reduction in NHEJ activity, and enhanced HDR efficiency (Canny et al., 2018). Moreover, it has been demonstrated that when the DNA replication inhibitor Geminin is fused with the N-terminus of Cas9, Cas9 activity is restricted to the S/G2/M phases of the cell cycle thereby promoting the probability of HDR (Gutschner et al., 2016).

The predominant DSBR pathways discussed above are schematically depicted in **Figure 4.1**.

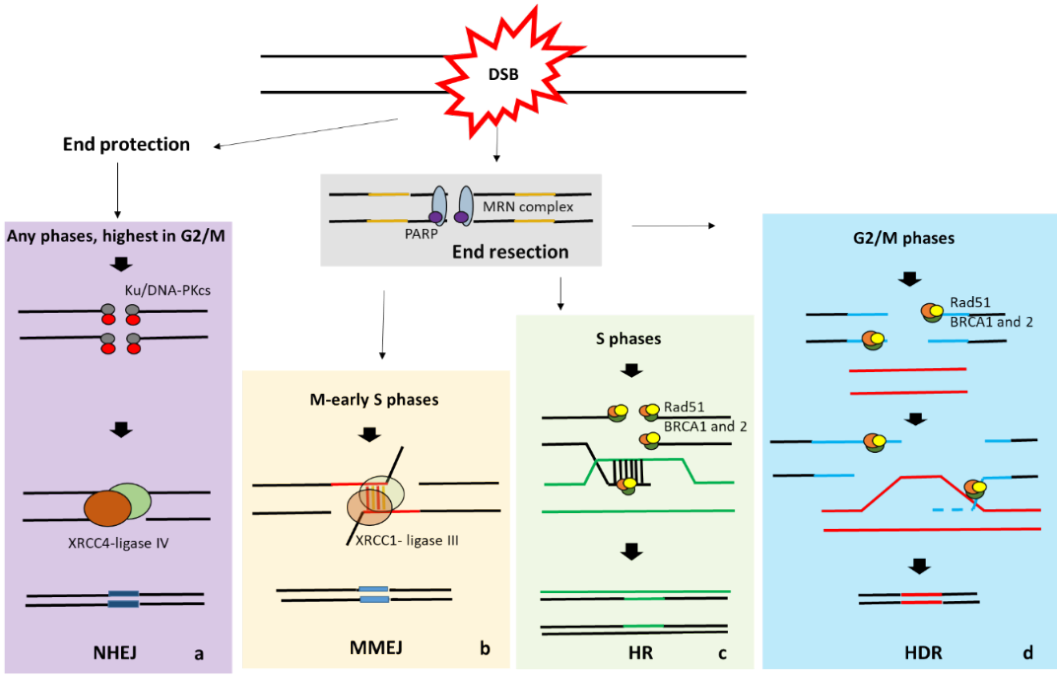


Figure 4. 1 Schematic presentation of dominant models of DSBR. During c-NHEJ repair of a DSB (a), the Ku70-Ku80/DNA-PKcs heterodimer binds to DNA ends, followed by DNA resection, synthesis and ligation, often resulting in small (1-4 bp) in/dels (dark blue box). Other repair pathways actively prevent end resection; micro-homologies are recognised by the MRN complex (light blue oval) and PARP1 (purple circle), inhibiting DNA resection. In MMEJ (b), the overhangs are removed, and the gap is filled and ligated by DNA ligase III-XRCC1, resulting in deletions of variable sizes (blue box). In HR (c) and HDR (d), a DNA tail exposed after DNA resection is bound by the Rad 51 complex, forming a homologous template for strand invasion. Subsequently, dsDNA is ligated and Holliday junctions are resolved. The cell cycle stages during which each repair mechanism predominantly functions are indicated above each scheme. NHEJ is active during all cell cycle stages and highest during G2/M. HR is most active in S phase, with a reduced activity during G2/M phases when HDR is most active (Dominguez et al., 2016).

As mentioned above, NHEJ and HR pathways are alternative pathways for the repair of DNA resections. The probability of repair by each repair pathway is associated with a wide range of factors, such as the species of organism, cell type, cell cycle progression, the presence of appropriate template molecules for HR and the level of DSB end resection. Because of the differences between the two central repair pathways, a diverse set of genome manipulation tools have been developed to precisely engineer the genome at the site of DNA repair.

	NHEJ	HR
Donor template	No requirement for templates	Template DNA molecule required (ssODN/plasmid)
Fidelity of repair	Low	High
Initial DNA end processing	No/limited processing of ends	Resection of 5' strands to form 3' tails.
Predominant cell cycle stage	Any phases, although repair during G1 predominates	S and G2 phases (Heyer et al., 2010)
Efficiency and precision	high but error-prone	Low but fewer errors

Table 4. 1 The critical bifurcations of NHEJ and HDR events. The summary is based on the report by Shuren Liao (Liao et al., 2016).

4.1.2 Genome manipulation

4.1.2.1 Animal transgenesis (TG)

Selective animal breeding has been widely used to promote favourable traits in animal herds. While there are several advantages to this approach, only half of the alleles pass from each parent to their offspring, so traits requiring a homozygous complement can skip generations. Although selective breeding is used in many agricultural systems, it can also result in inbreeding, reduced genetic diversity and potentially perpetuate negative co-segregating traits. Furthermore, selective breeding can only be based upon the existing genetic characteristics present in the animal herds. In order to introduce specific novel traits in animal breeding, several strategies have been applied to animal breeding to generate genetically modified organisms (GMOs).

GMOs are the organisms whose genome has been purposefully modified using genetic engineering techniques. GMO includes the alteration or deletion of endogenous genes. The first GMO was exemplified in 1972, when a constructed plasmid encoding a toxin gene was transformed into the bacterium *Escherichia coli*, and inserted into the bacterial genome (Stanley N. Cohen, 1972). The first GM birds were produced using a retroviral vector delivery system in 1997. Moreover, GMO includes the introduction of exogenous genes. The term transgenesis (TG) was first used in 1981 to describe pronuclear injection (PI) of SV40 DNA into the zygotes of mice to produce animals harbouring a foreign DNA (Gordon and Ruddle, 1981). Subsequently, the techniques was applied to livestock. There are some limitations, however, including random insertional mutagenesis in the genome, the inefficiency of electroporation, high efforts in time and money. Other viral delivery systems such as retrovirus, adeno-associated virus (AAV) and Herpes simplex virus (HSV) were

developed to transfer the foreign transgene to an organism. Some sequential landmarks of GM/TG development over the 46 years since the 1972 report are shown**:**

Year	Milestones of GMO/TG	Reference
1972	First report of GMOs- bacteria *Escherichia coli*	(Stanley N. Cohen, 1972)
1974	First chimeric mouse by blastocyst microinjection	(Jaenisch and Mintz, 1974)
1975	First retrovirus-mediated nucleic acids delivery	(Jaenisch et al., 1975)
1981	First pronuclear injection (PI) of foreign DNA in mouse	(Gordon et al., 1980)
1982	First GM insect with germ line transformation	(Rubin and Spradling, 1982)
1983	First GM plants with Agrobacterium tumefaciens	(Bevan et al., 1983)
1985	First PI of foreign genes in pigs and sheep	(Hammer et al., 1985b)
1986	Embryonic cloning by NT in sheep	(Willadsen, 1986)
1989	First GM birds with retroviral vectors	(Bosselman et al., 1989)
1991	First TG sheep producing altered milk	(Wright et al., 1991)
1991	First TG cattle	(Krimpenfort et al., 1991)
1992	First cloned pig	(Fischer, 1992)
1992	TG pigs in resisting IAV infection	(Muller et al., 1992)
1997	First mammal cloned from somatic cell (Dolly)	(Campbell et al., 1996)
1997	TG livestock for modelling retinitis pigmentosa (RP)	(Petters et al., 1997)
1997	First NT associated with *in vitro* cell cultured sheep	(Schnieke et al., 1997)
2002	Alpha 1,3-Gal KO cloned pigs (SCNT)	(Dai et al., 2002)

2005	TG cow in resisting mastitis infection	(Wall et al., 2005)
2008	CFTR-KO/deltaF508 pigs	(Rogers et al., 2008)
2011	GGTA1 biallelic KO pigs (ZFNs)	(Hauschild et al., 2011)
2011	eGFP KO pigs (ZFNs)	(Whyte and Prather, 2011)
2012	LDLR KO pigs (TALENs)	(Carlson et al., 2012)
2013	RELA pigs (ZFNs and TALENs)	(Lillico et al., 2013)
2014	Mx1 overpressing TG pigs	(Yan et al., 2014a)
2014	Rosa26-targeted pig model (TALENs)	(Li et al., 2014)
2014	MSTN KO pigs (CRISPR/Cas9)	(Paquet et al., 2016)
2017	RPPSV resistance pigs (CRISPR/Cas9)	(Christine Burkard, 2017)
2017	PERVs inactivated pigs (CRISPR/Cas9)	(Niu et al., 2017)
2018	Huntingtin KI pig models PERVs inactivated pigs (CRISPR/Cas9)	(Yan et al., 2018)
2019	TGEV and PEDV resistant pigs PERVs inactivated pigs (CRISPR/Cas9)	(Whitworth et al., 2019)

Table 4. 2 Significant landmarks of GM/TG development. SCNT: somatic cell nuclear transfer; TG: transgenic; CSF: classic swine fever; CFTR: Cystic fibrosis transmembrane conductance regulator; GGTA: porcine alpha 1,3-galactosyltransferase; LDLR: low-density lipoprotein receptor; Mx1: myxovirus resistance gene; MSTN: myostatin; PRRPv: Porcine reproductive and respiratory syndrome virus; PERV: Porcine endogenous retrovirus; TGEV: gastroenteritis virus; PEDV: Porcine epidemic diarrhea virus.

Current genome manipulation methods still require improvements in terms of efficiency, time and cost. To enhance gene targeting by accurately manipulating DNA repair pathways, more efficient methods involving genome editing (GE) have been developed and discussed below.

4.1.2.2 Genome editing (GE)

Genome editing (GE) refers to a process in which the genome of living organisms is altered in a deliberate manner. Many of the current genome editor reagents create a nuclease-induced DSB to cut in the genome proximal to a site of interest with subsequent repair either creating in/dels or including precise modifications to the DNA sequence. GE has shown its powerful utility in basic scientific research and industrial applications.

4.1.2.3 FokI

FokI, a type-IIS restriction endonuclease isolated from *Flavobacterium okeanokoites* (Chandrasegaran and Smith, 1999), recognizes the non-palindromic nucleotide sequence 5′-GGATG(N)$_{9/13}$-3′ (Sugisaki H et al., 1981). There are two separate domains to FokI, a DNA recognition domain and a catalytic cleavage domain. The latter domain lacks sequence specificity and can be further split into two subunits that require dimerization to be functional (Dorner et al., 1999).

4.1.2.4 Zinc-Finger (ZF) and ZF-nuclease (ZFN)

Miller described a short (~30 amino acids) DNA-binding motif, which consisted of nine repeated units in the transcription factor TFIIIA of *Xenopus laevis* (Miller et al., 1985). Each unit contains one α–helix and two β-sheets with an oriented

direction, and a central zinc ion (Pavletich and Pabo, 1991). These motifs formed a single protein structure, the zinc finger (ZF), which specifically recoginises three successive DNA bases, with the α–helix binding to the major groove of the target DNA. As these repeats harboured a high frequency of cysteine (Cys) and histidine (His) residues, they are referred to as the Cys2-His2-ZF proteins. By fusing several fingers into an array, it became possible to expand the recognition sequence of a synthetic ZF protein, providing a high degree of DNA binding specificity.

The endonuclease domain of FokI can be fused to a synthetic ZF protein by means of a spacer to produce a chimeric restriction enzyme, the zinc finger nuclease (ZFN), allowing scientists to design *de novo* nucleases with high sequence specificity for a target sequence shown in **Figure 4. 2**. Typically, ZFNs are deployed as a pair proteins, each with specificity for a 9 bp genomic sequence and fused to half of the FokI deterodimer. Each half of the pair binds in an inverted position to the complementary strands of DNA, binding the two halves of the FokI into position to dimerise. A DSB was thereby created. ZFNs have been employed in research in numerous species including human (Gaj et al., 2013), zebrafish (Doyon et al., 2008), pig (Hauschild et al., 2011), rat (Geurts et al., 2009), C.elegans (Wood et al., 2011) and drosophila (Bibikova et al., 2002) to induce a gene disruption; and in human (Hockemeyer et al., 2011) and mouse (Perez-Pinera et al., 2012) to facilitate a site-specific integration of a transgene. For example, CCR5, as a receptor of HIV-1, has been deleted in CD4+T lymphocytes by ZF nucleases, providing robust approaches in HIV resistance (Perez et al., 2008); pigs with a biallelic KO of an endogenous and immune-related gene, α 1,3-galactosyltransferase (GGTA1), were created. This could be applied in the generation of pig KO strains, and the generations of the novel

therapies in biomedicine (Hauschild et al., 2011).

However, the ZFNs system did not have huge uptake by the scientific community, predominantly because sequential ZF domains often have crosslinks that can influence the sequence specificity (Cornu et al., 2008). Therefore, many designed ZFNs failed to work as intended, and pricing for commercial design and synthesis was often prohibitive.

4.1.2.5 TALE and TALEN

Transcription activator-like effectors (TALEs), secreted by the from *Xanthomonas* bacteria, bind specific promoter sequences in the host and upregulate genes that enhance bacterial infections (Römer et al. 2009). The basic unit structure of the TALE protein is a 33-35 monomer, 2 amino acids of which confer specificity for binding a single DNA base. The array of amino acid repeats and the DNA nucleotide sequence interact in a one-to-one recognition mode. This recognition pattern makes TALEs ideal for binding DNA loci. Like ZFs, a TALE array can be fused with a FokI endonuclease to create a transcription activator-like effector nuclease (TALEN) (Christian et al., 2010). TALENs are typically employed in pairs, targeting opposing DNA strands to allow dimerisation of FokI as shown in **Figure 4. 2b**.

TALENs are easier to construct and engineer than ZFNs. However, TALENs are approximately three times the length of ZFNs, meaning that their cDNA is not compatible with some vector systems. This results in more labour, and difficulties in transfer from *in vitro* studies to *in vivo* application such as biomedical therapies (Holkers et al. 2013). In addition, TALENs were initially reported to require a 5’

thymidine (T) in the recognition sequence with deviation resulting in significantly reduced binding efficiency (Lamb et al. 2013).

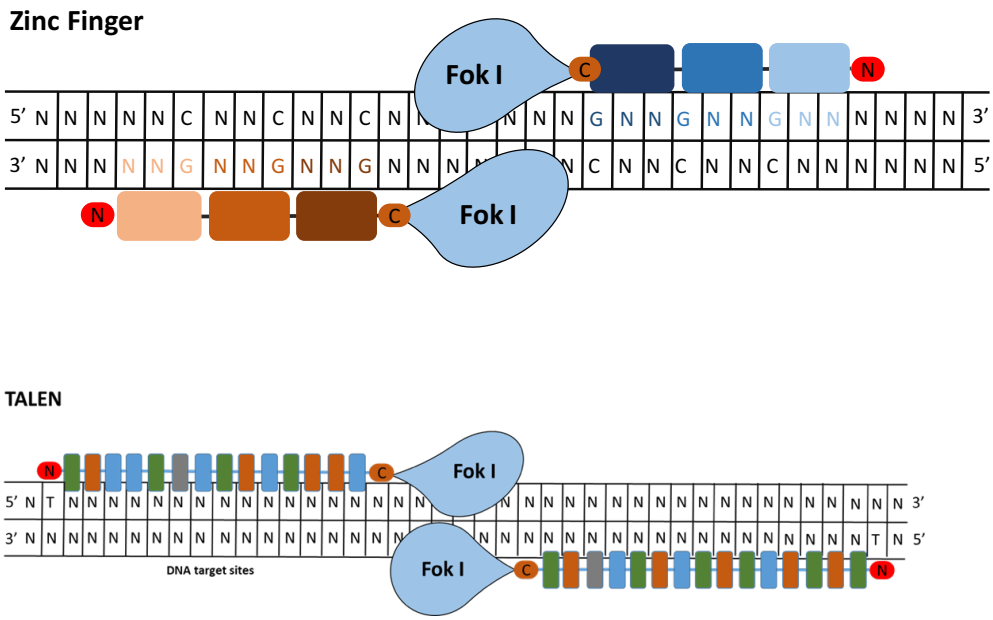


Figure 4. 2 ZFN and TALEN structures. The DNA binding domain of ZF (a) or TALE (b) can be fused to the catalytic domain of FokI endonuclease to induce a DSB in the genome. DNA-binding domains and FokI catalytic domain work in pairs and locate in the opposite of DNA double strands separated by a spacer sequence. ZF motif binds to DNA triplets (gradient orange and blue boxes), while each TALE repeat binds to a single nucleotide (coloured boxes).

4.1.2.6 CRISPR and CRISPR/Cas9 system

In 1987, small DNA repeated motifs separated by a DNA spacer were found in *E. coli* (Aasheim et al., 1993), and were subsequently coined 'Clustered Regularly Interspaced Short Palindromic Repeats' (CRISPR) (Diez-Villasenor et al., 2010). The functions of these repeats remained unclear until Barrangou revealed CRISPR provided acquired resistance against foreign DNA in prokaryotes and characterised Archaea (Jansen et al., 2002). a 20 bp fragment of DNA from an invading pathogen is inserted into the bacterial genome and transcribed into a target-specific CRISPR RNA (crRNA). crRNA hybridises with a second, target-independent trans-activating crRNA (tracrRNA) (Deltcheva et al., 2011), to form a complex, which in turn complexes with by bacterial CRISPR-associated (Cas) protein. If the bacterial cell reencounters the same foreign DNA, the 20 bp crRNA can hybridise with the complementary invading sequence, targeting the Cas protein to cut the DNA proximal to a short protospacer adjacent motif (PAM) (**Figure 4. 3**).

There are two classes of the Cas proteins have been characterised to date, class one (type I, III, IV) and class two (type II, V, VI) (Seed et al. 2013). The class one system employs a multi-protein complex while the class two system uses a single effector protein to mediate the cleavage of target DNA. As the class two system is more amenable to manipulation, it has become the most utilised system for the tool design. A wealth of class two Cas proteins have now been identified from a range of bacterial species. *Streptococcus pyogenes* cas9 (*Sp*Cas9) protein, has been the most widely utilised. A CRISPR RNA (crRNA), which encodes the guide RNA by base pairing, and a trans-activating RNA (tracrRNA), which is a endogenous component (Garneau et al., 2010), can be fused to create a chimeric, single-guided RNA (sgRNA).

The sgRNA and *Sp*Cas9 complex can bind to DNA complementary to the sgRNA and 5' to a PAM (5'-NGG-3'). After this, two distinct nuclease domains in *Sp*Cas9, HNH and RuvC, cleave complementary and non-complementary domain of DNA strands, respectively. The cleavage occurs at approximately 3–5 bp upstream of the PAM. This blunt end produced will be repaired by one of the DSBR mechanisms introduced previously (**section 4.1**).

So far, the rate of both technological advances and community uptake of CRISPR/Cas9 system-mediated editing has been stunning and is now broadly applied in basic biological research and biomedicine (Eid and Mahfouz, 2016). Cas9 has been applied in different organisms, such as human cells (Pattanayak et al., 2013), mice (Cong et al., 2013), rats (Doyon et al., 2008), rabbits (Yan et al., 2014b), fruit flies (Awata et al., 2015)and plants (Shan et al., 2013).Very importantly, edited pigs have been made for a wide range of applications using the CRISPR/Cas9 system: first genome-modified pig (Hai et al., 2014), MSTN KO pigs (Bi et al., 2016), PRRSV resistant pigs (Christine Burkard, 2017), PERVs inactivated pigs (Niu et al., 2017), UCP1 KI pigs (Zheng et al., 2017), Huntingtin KI pig models (Yan et al., 2018), IGF2 edited pigs (Xiang et al., 2018), TGEV and PEDV resistant pigs (Whitworth et al., 2019).

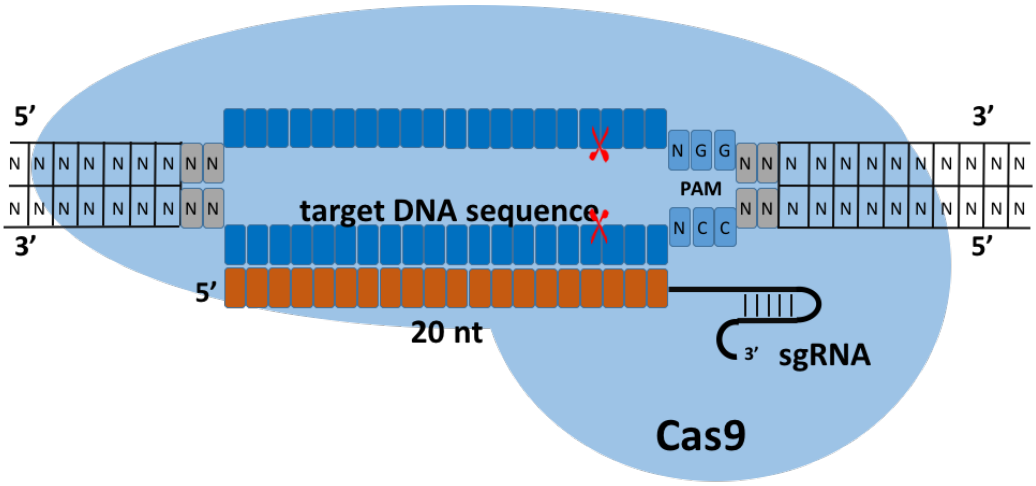


Figure 4. 3 Schematic representation of the CRISPR/Cas9 complex. In the CRISPR/Cas9 system, a single guide RNA (the complex contains crRNA and tracrRNA) recognises and binds to a specific sequence on the genome (blue box). It guides Cas9 endonuclease enzyme (blue) to target and cleave the genome. This 20 nt targeted locus of DNA is complementary to the chimeric sgRNA (orange box).

4.2 Strategy overview

4.2.1 Editing the transcriptional profiling of pig ST6GAL1

In the first strategy, we used CRISPR/Cas9-based system to target ST6GAL1 mRNA transcripts predominantly expressed in the respiratory tract, which were identified experimentally in **Chapter 3**. We anticipated that the alteration of ST6GAL1 transcriptional profile would render trachea cells deficient in using the 5' transcription start site (TSS) to initiate the transcription in a promoter-mediated manner. To this end, two CRISPR gRNAs targeting 500 bp upstream E (-5) and 500 bp downstream E(-4) were designed. This was because DNA sequence involving 200 bp upstream of the 5' TSS is defined was defined a promoter region (Pieter Meysman, 2014). Dual sgRNAs were designed to remove the interval (~2.9 kb region) between them, resulting in a model termed as '-/-ΔP' (**Figure 4. 4**).

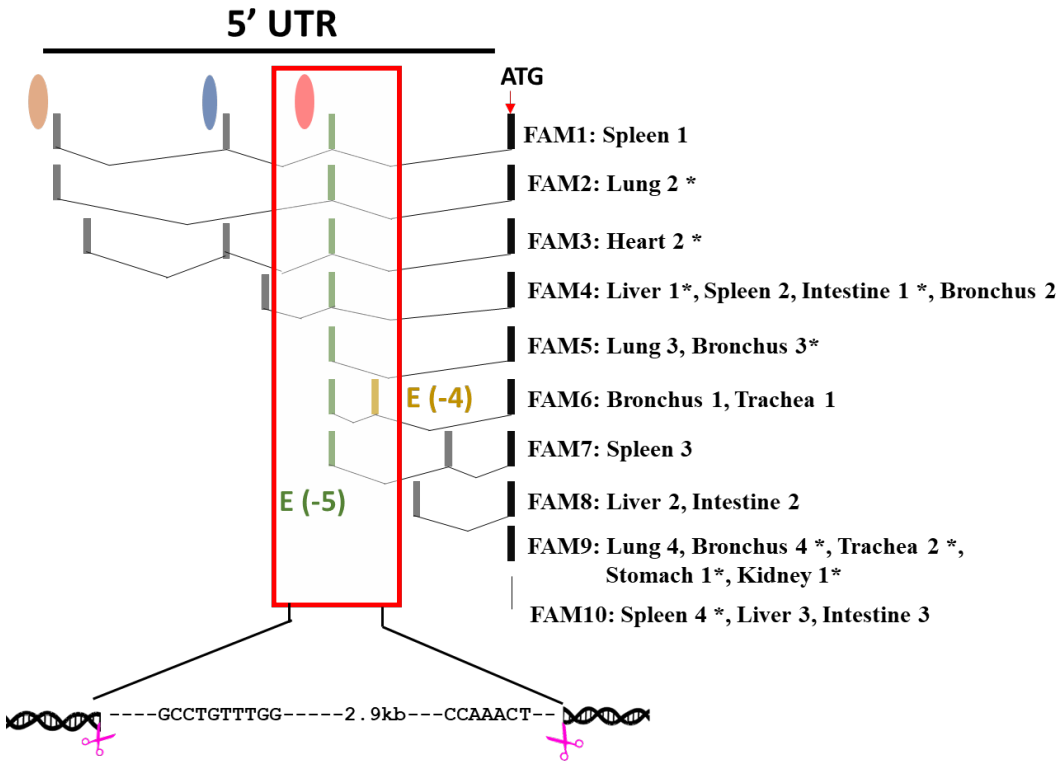


Figure 4. 4 5' UTR region of pig ST6GAL1 transcripts. The target region is in the dashed pink box. Double sgRNAs (pink scissors) were designed spanning untranslated E(-4) and E(-5), and an associated predicted promoter region (the light red circle). All transcripts enriched in respiratory tracts are labelled on the right. UTEs are denoted by grey boxes, and the first coding exon is the black box.

4.2.2 Functional knockout of pig ST6GAL1

In the second strategy, a CRISPR guide was designed downstream of the third AUG in the first coding exon of the ST6GAL1 gene. The target site was accordingly located between the co-localization domain (CLD) and the catalytic domain (CAD) of the protein (**Figure 4. 5**). It was anticipated that frameshift mutation within coding sequence might be introduced, and thus the translation of protein could be influenced.

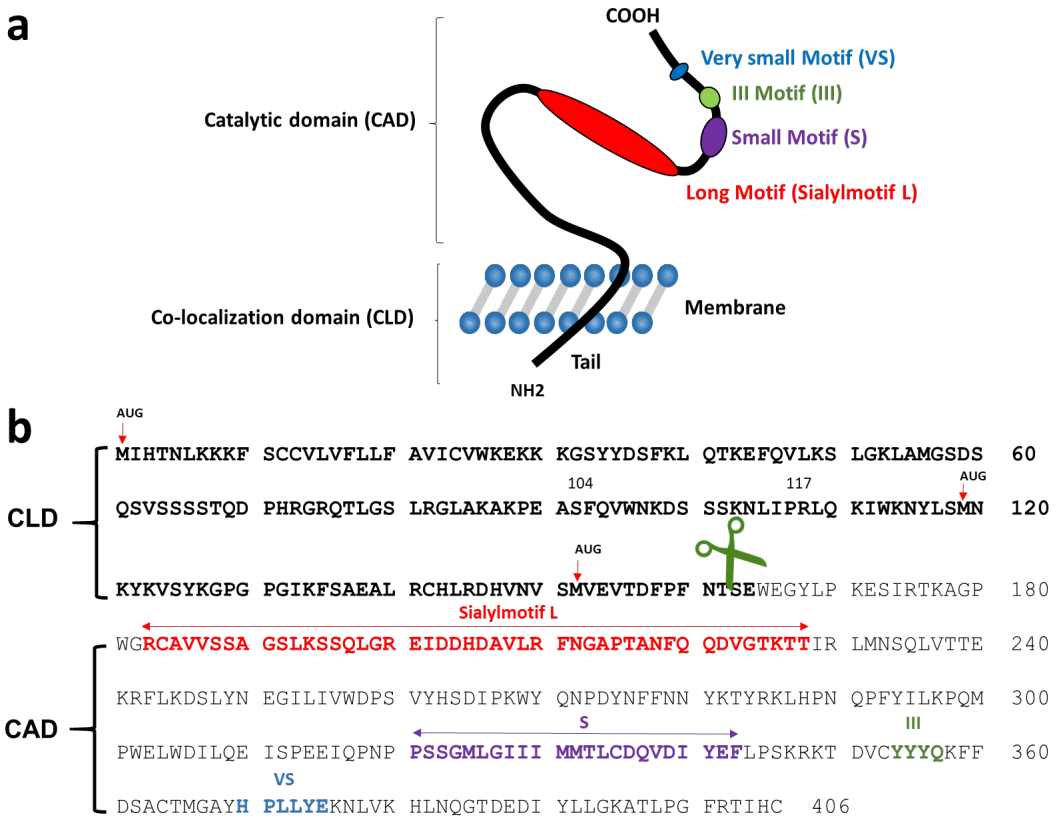


Figure 4. 5 ST6Gal1 protein. (a) The structure of the ST6Gal1 protein, with the catalytic domain (CAD) and the co-localization domain (CLD) highlighted. The conserved and functional peptides are L, S, and VS sialylmotifs. L and S motifs are important in site-directed mutagenesis. Specifically, L-sialymotif is associated with donor substrate binding; S-sialymotif is with both donor and acceptor binding. (b) CRISPR gRNA (green scissor) was designed accordingly downstream the third AUG.

4.3 Result 1—precisely engineering ST6GAL1

4.3.1 CRISPR/Cas9 reagent construction

Candidate sgRNAs with high specificity and minimal predicted off-target Cas9 binding activities were selected as described previously (**Section 2.3.1.9**). Each sgRNA was then inserted into a Cas9 expression vector, pSL66 (map in **Appendix 7**), which has been successfully used for gene editing in pig cells (Christine Burkard, 2017). The vector contained a designed gRNA sequence which was driven by a human U6 (hU6) promoter, as well as the *Sp*Cas9-2A-GFP sequence which was driven by a chicken β-actin promoter (CBh) promoter (**Figure 4. 5-a**).

Oligonucleotides encoding sgRNAs were annealed and cloned into the pSL66 vector via *BbsI* restriction sites with resultant expression plasmids being assigned an 'oDY' number. The visualization of bacterial PCR products on a 2% agarose gel exhibited correct PCR fragments amplified (about 360 bp) (**Figure 4. 5-b**). Each PCR fragment was then cut out, and purified for Sanger sequencing (**Figure 4. 5-c**).

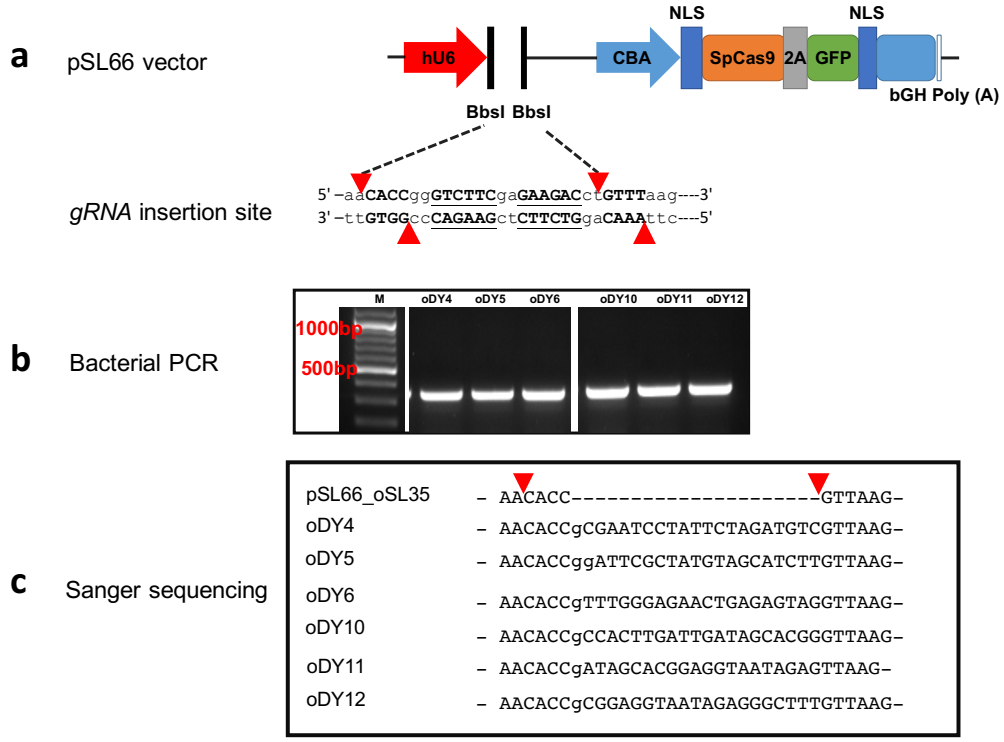


Figure 4. 6 Successful CRISPR assembly. (a) Schematic representation shows that the sequence on the cloning sites was inserted by the annealed oligos such that sgRNA expression would function from the hU6 promoter in pSL66 vector; the gRNA sequence was inserted downstream hU6 promoter via *BbsI* sites. (b) The colony PCR result exhibited an example of PCR products resolved on a 1% TAE gel; (c) The alignment of Sanger sequencing analysis of colony PCR products, with unmodified pSL66 (pSL_U6_oSL35) as reference. hU6-human U6 promoter; CBh-chicken β-actin promoter; NLS-nuclear localization signal; bGH-bovine growth hormone.

4.3.2 Transfection and CRISPR guides validation

4.3.2.1 Transfection efficiency in PK15 cells

We first tested the transfection efficiency of sgRNA in porcine kidney cell line (PK15) cells by electroporation. Briefly, we transfected 1 µg CRISPR/Cas9-GFP plasmid in ~5 x 10^5 PK15 cells with the optimised conditions using a Neon™ electroporation system, i.e., 1400 V (pulse voltage), 30 ms (pulse width), and one electrical pulse. As shown in **Figure 4. 7a,** the mock-transfection, in which PK15 cells were not transfected (-ve), exhibited no fluorescent signals. Positive control was a sample transfected with the pSL66 vector without inserted sgRNA sequence (+ve), exhibiting fluorescent GFP signals (green). All the transfected cells showed GFP expression, indicating successful transfection.

At 24 hour post transfection (h.p.t), positive cells were enriched by fluorescent activated cell sorting (FACs), based on the GFP fluorescent signals (**Section 2.1.1.5**). Forward scatter (FSC) and side scatter (SSC) properties were used to gate the healthy cells and the single cell events. The quantititive analysis of GFP expression in positive control is shown in **Figure 4. 7b** as an example. The cell viability at 24 h.p.t was 69%, compared to that from mock-transfected cells. Transfection efficiency was then determined as 2.8%, which was the percentage of GFP+ cells in a transfected cell population. Similarly, the transfection efficiency with oDY4, 5 and 6 were 2.4%, 0.8%, and 0.7%, respectively; the transfection efficiency with oDY10, 11 and 12 were 5.2%, 3.0% and 3.8%, respectively. All the flow cytometry results are shown in **Appendix 8.**

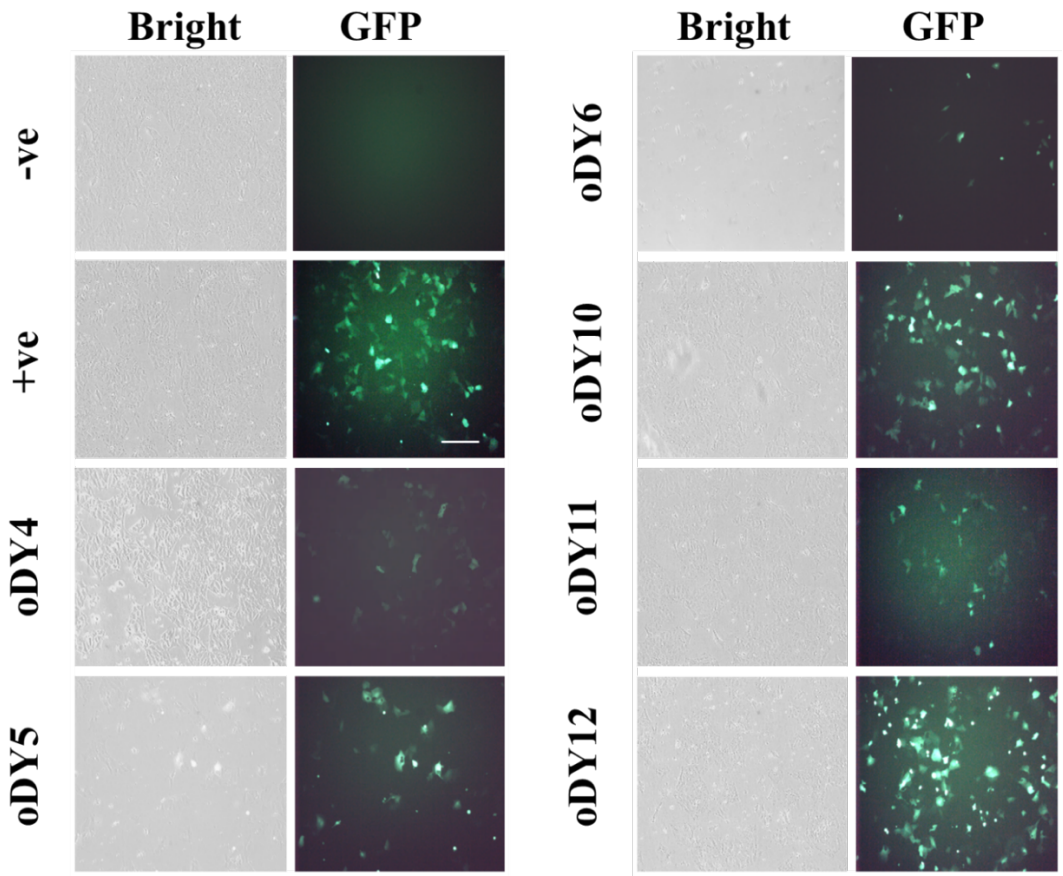


Figure 4. 7 DNA transfection using the Neon™ electroporation system and flow cytometry analysis. (a) PK15 cells were transfected with CRISPR/Cas9-GFP plasmid with Neon™ transfection system. Scale bar=100 µm

4.3.2.2 Evaluating the cutting efficiency of sgRNA using T7E1 assay

A T7 endonuclease 1 (T7E1) assay was performed to evaluate the on-target CRISPR/Cas9 editing events (Vouillot et al., 2015). Firstly, a pair of primers was designed to amplify a 500-1000 bp DNA sequence surrounding the CRISPR gRNA target site. Prior to PCR amplification, the annealing temperature of primers was determined across a gradient of 12 temperatures ranging from 50˚C to 72˚C. The optimal annealing temperature was determined when a single robust amplicon was produced on a 2% agarose gel. The result of gradient PCR is shown in **Appendix 9**. 68˚C was determined to be the optimal annealing temperature for primer pair 2 (for sgRNA4, 5, and 6), and 65˚C for primer pair 7 (for sgRNA10, 11, and 12).

Using the optimised PCR conditions, PCR was performed with gDNA isolated from each GFP+ cell population. PCR products were then denatured and re-annealed, followed by a mock digestion and T7E1 digestion. Mock-digested and T7E1-digested products were resolved on a 1% agarose gel, as shown in **Figure 4. 8**.

For sgRNA4, 5, and 6, primer pair 2 (F2 and R2) was used to amplify a DNA sequence (964 bp) flanking the DSB site (**Figure 4. 8-a**). Compared with the mock-digested product, each T7E1-digested product presented two successful cleavage of DNA, which were labelled below the gel (**Figure 4. 8-b)**. sgRNA4 had 40.22% cutting efficiency, and sgRNA5 had 41.01% cutting efficiency. sgRNA6 exhibited the highest cutting activity (47.97%). It was noted that all T7E1-digested PCR products had an extra band of ~700 bp. This indicates the probable presence of single nucleotide polymorphisms (SNP) within the amplicon, either present within the genome of the PK15 cells or introduced during PCR amplification. Given that this fragment was

present in each experimental sample, it seems likely that this SNP was a feature of the PK15 cells. Therefore, this band was excluded from the densitometry analysis in DNA cleavage.

The same procedure was performed on the gDNA isolated from GFP+ cells individually transfected with oDY10, 11 and 12 (**Figure 4. 8-d)**. The PCR performed using primer F7 and R7 resulted in a PCR product with a size of 792 bp. The cutting efficiency of sgRNA11 and 12 was estimated at 13.2% and 33.23%, respectively; sgRNA10 showed no evidence of cutting. Therefore, sgRNA6 and 12, which exhibited relative higher cutting efficiencies than the other sgRNA candidates, were co-transfected in PK15 cells.

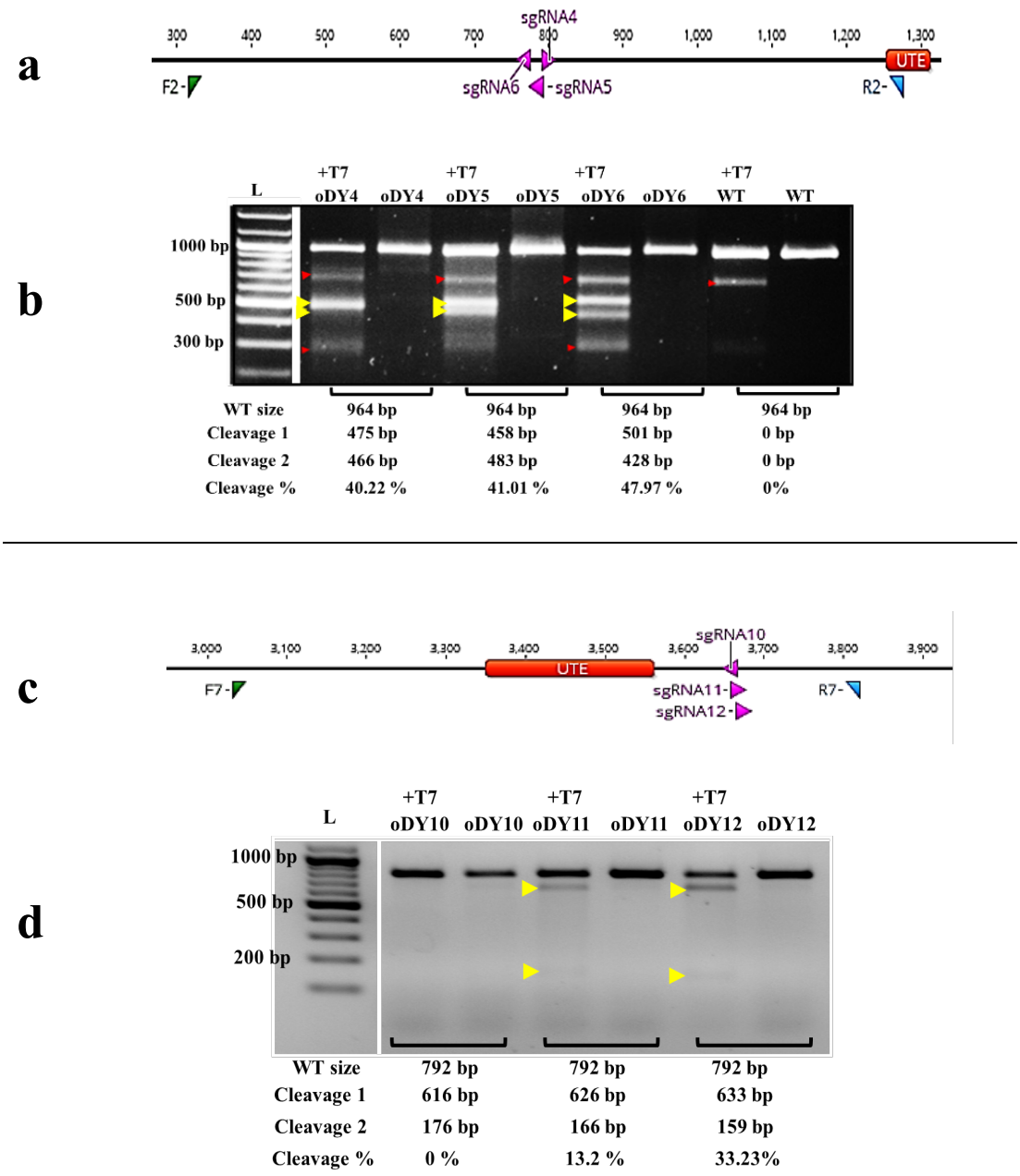


Figure 4. 8 T7E1 assay. (a)Diagram shows the schematic diagram of PCR surrounding the target site of sgRNA4, 5 and 6 using primer pair 2 (964 bp). (b)Expected cleaved products of digested PCR products. Yellow arrows indicated the anticipated DNA cleavage in each T7E1-digested product. Red arrows indicated the extra bands. (c) PCR surrounding the target site of sgRNA10, 11, and 12 using primer pair 7 (792 bp). (d) Expected cleaved products of digested PCR products. Yellow arrows indicated the anticipated DNA cleavage in each T7E1-digested product. The cutting efficiency of each sgRNA was below the gel. The band intensity on the gel was measured by ImageJ. CRISPR/sgRNA-GFP plasmid was assigned as 'oDY'. L - 1 kb ladder.

4.3.2.3 Co-transfection with dual CRISPR guides in PK15 cells

Co-transfection of Cas9 expression plasmids encoding sgRNA6 and sgRNA12 were performed to remove the interval between sgRNA6 and sgRNA12 in PK15 cells. GFP+ cells were enriched at 24 h.p.t by FCSs, followed by the measurement of transfection efficiency (see **Figure 4. 9)**: the mock-transfection (-ve) exhibited no GFP fluorescent signals. The positive control (+ve), in which PK15 cells were transfected with 2 µL pSL66 vector without a guide sequence, had a 13.3% transfection efficiency. Co-transfection with 1 µL oDY6 and 1 µL oDY12 yielded a transfection efficiency of 13.8%.

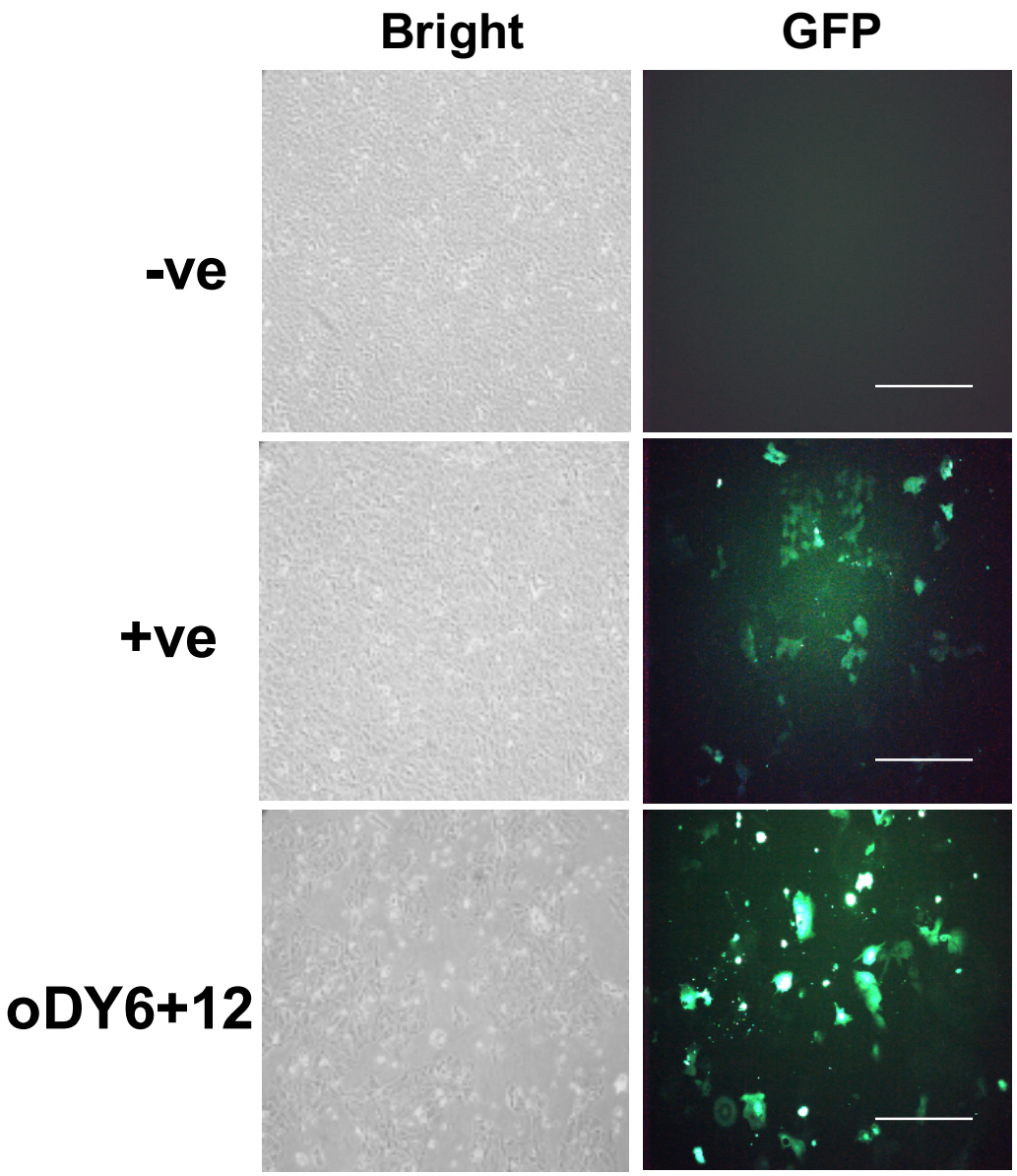


Figure 4. 9 Transfection efficiency of dual sgRNAs in PK15 cells. Co-transfection dual sgRNAs/Cas9 plasmids oDY6 and oDY12 in PK15 cells by Neon™ electroporation. Scale bar=100 μm.

gDNAs were then isolated from WT cells and GFP+ cells for PCR amplifications. The PCR amplification using primers CoF1 and CoR1 was anticipated to result in a ~3.5 kb product in WT, and a ~0.6 kb product due to the absence of the 2.9 kb target region in the DNA sequence of edited cell (**Figure 4. 10a**) The results are shown in **Figure 4. 10b**: there appeared no bands in WT, due to the suboptimal efficiency in amplifying a large WT fragment. The PCR result of GPF+ cells exhibited a correct cleaved product (~0.6 kb), indicating the successful deletion mediated by two CRISPR guides.

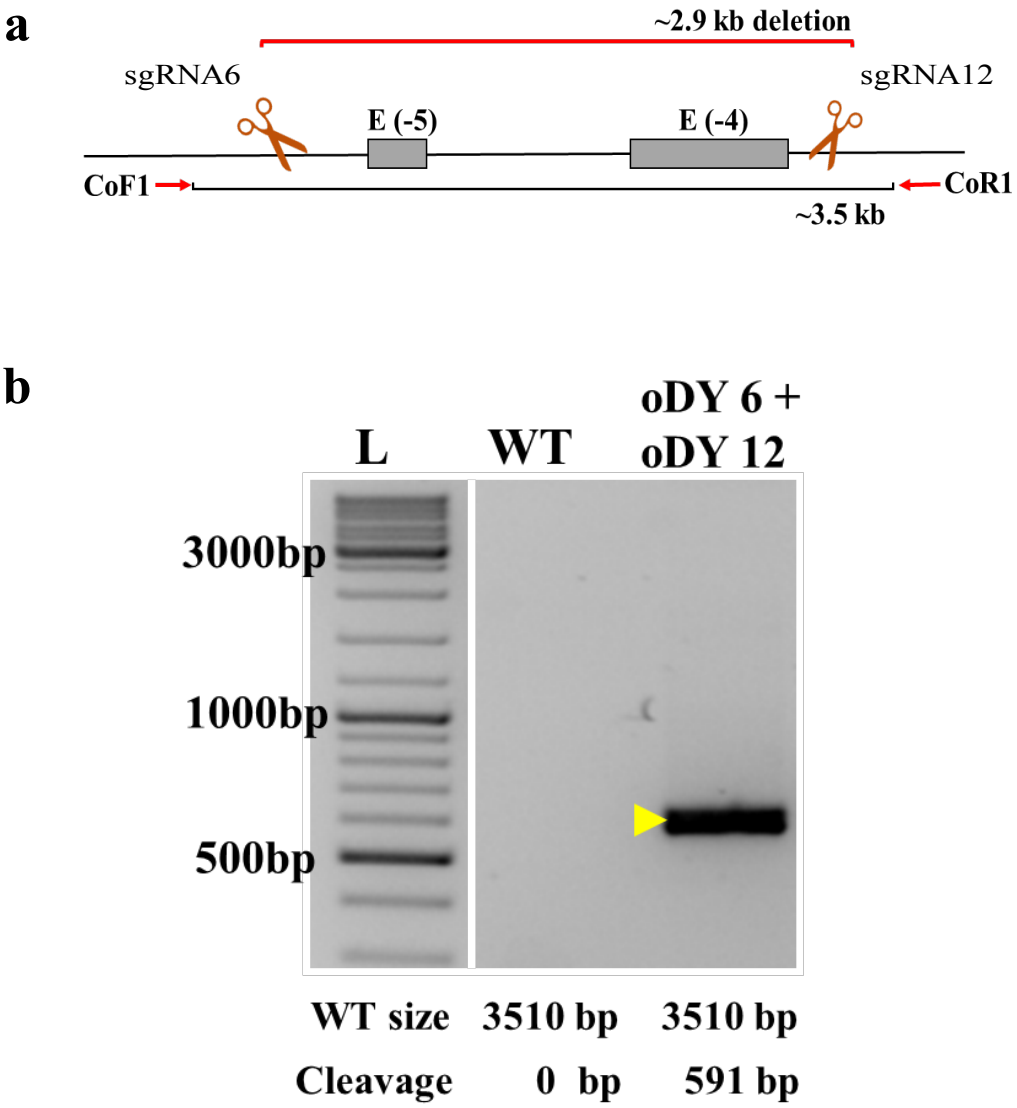


Figure 4. 10 Genetic deletions with dual sgRNAs. (a) Schematic overview of the guide sites and PCR assay for assessing the genetic deletion. Two PCR primers (red arrows), spanning the target region, were used to amplify a ~3.5 kb DNA fragment in WT cells and a ~2.9 kb product in edited cells. (b) PCR products were resolved on a 2% agarose gel. The anticipated size of the WT and deletion amplicons were shown below the gel (annealing temperature was 68°C).

4.3.2.4 Assessment of CRISPR guides in NPTr cells

PK15 cells were a good starting point used in the optimisation of the transfection protocols and testing CRISPR reagents in PK15 cells. Since the aim of the project was to assess the IAV infection in ST6GAL1 edited cells, we performed the similar transfection protocol in a newborn pig trachea (NPTr) cell line, which was a proper cell model for IAV infection. We optimised conditions of Neon^{TM} transfection by varying the concentration of Cas9/sgRNA plasmid, the number of cells, and the electric pulse in NPTr cells. However, the performance of Neon^{TM} transfection at 24 h.p.t was poor in NPTr cells despite adjusted parameters (**Figure 4. 11-a)**. For example, compared to that in A2, the number of live cells in B6 was reduced with the increasing level of the electricity pulse and the concentration of the plasmid DNA (**Figure 4. 11-b**). In this situation, an appropriate transfection protocol for transfection in NPTr cells was important.

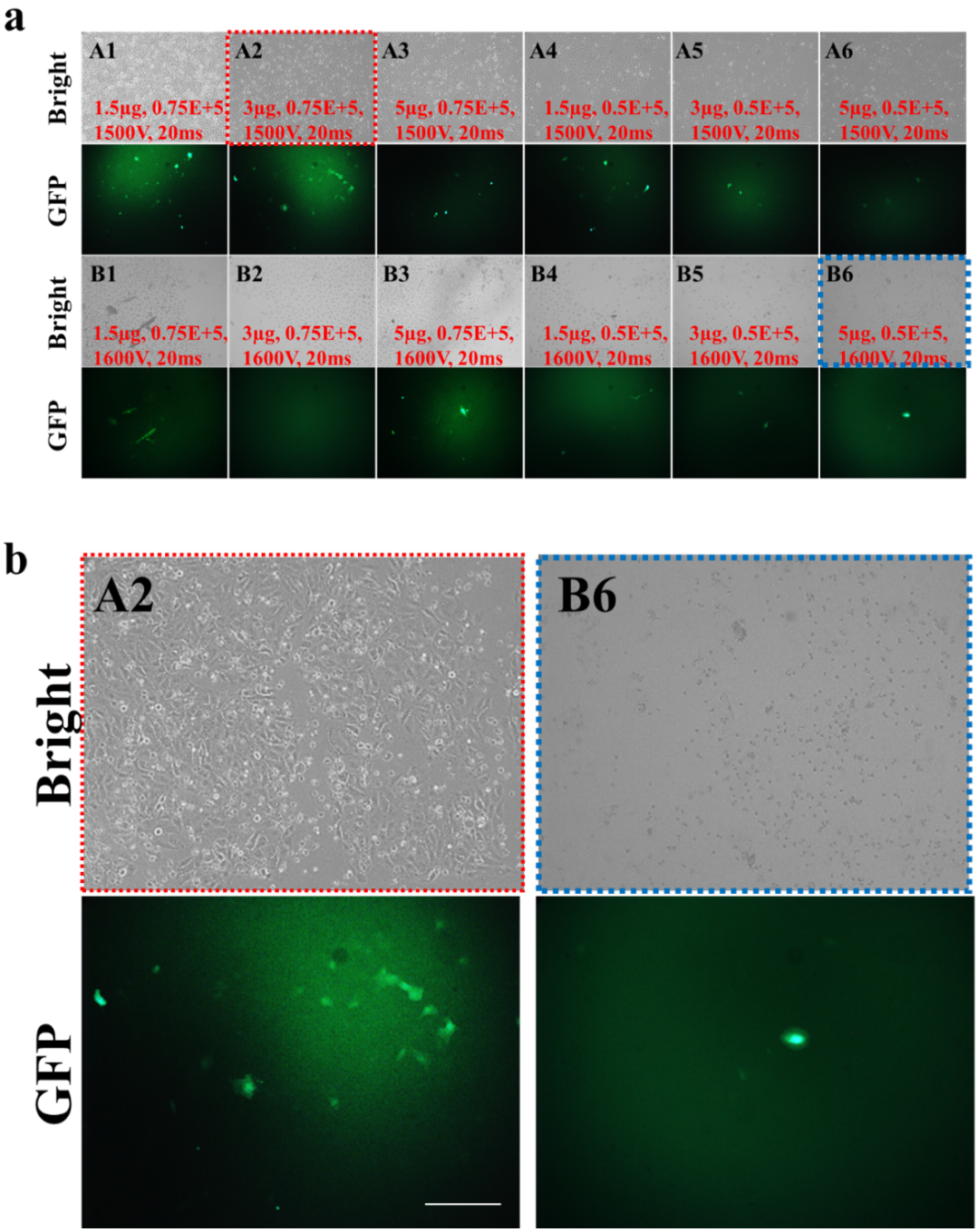


Figure 4. 11 Neon™ transfection optimization in NPTr cells. (a) Transfected cells at 24 h.p.t with conditions labelled in red. (b) The expanded views of A2 and B6 show the cell viability and GFP+signals. The parameters are listed here: A1-1.5 µg, 0.75E+5 cells, 1500V, 20ms; A2-3 µg, 0.75E+5 cells, 1500V, 20ms; A3- 5 µg, 0.75E+5 cells, 1500V, 20ms; A4- 1.5 µg, 0.5E+5 cells, 1500V, 20ms; A5- 3 µg, 0.75E+5 cells, 1500V, 20ms; A6- 5 µg, 0.75E+5 cells, 1500V, 20ms. B1-1.5 µg, 0.75E+5 cells, 1600V, 20ms; B2- 3 µg, 0.75E+5 cells, 1600V, 20ms; B3- 5 µg, 0.75E+5 cells, 1600V, 20ms; B4- 1.5 µg, 0. 5E+5 cells, 1600V, 20ms; B5- 3 µg, 0. 5E+5 cells, 1600V, 20ms; B6- 5µg, 0. 5E+5 cells, 1600V, 20ms. All samples were performed with one electric pulse. Scale bar=100 µm

Lipofectamine® 2000 system has been widely used in *in vitro* and *in vivo* mammalian cells transfection (Colosimo et al., 2000) with higher editing efficiency and less cell toxicity (Yu et al., 2016). Therefore, Cas9 protein and sgRNA-containing plasmids (oDY6 and oDY12) were transfected in NPTr cells using the Lipofectamine® 2000 system. The transfection was performed according to the manual protocol. Briefly, 15 µl Lipofectamine® 2000 reagent and 2 µg DNA (1 µg for each plasmid) were mixed in 300 µl, and the mixture was added to 5 x 10^5 cells in 6-well plates. At 24 h.p.t, transfected cells were visualised and quantitatively assessed (**Figure 4. 12-a**). The negative control showed no GFP+ fluorescence, and the positive control (+ve) exhibited 2.7% transfection efficiency. The cells with dual sgRNAs/CRISPR co-transfection exhibited 6.5% transfection efficiency.

To improve the transfection performance, we optimised the conditions empirically, i.e., 12.5 µl Lipofectamine® 2000 reagent and 3 µg plasmid (1.5 µg for oDY6 and 12 plasmid) were mixed in 300 µl transfection volume, and 200 µl mixture of the mixture was applied dropwise in about 2 x 10^5 NPTr cells. This approach resulted in a substantial improvement of the transfection efficiency (**Figure 4. 12-b**). The positive control gave 23.8% and co-transfected oDY6 and oDY12 gave 22.9% GFP cells at 24 h.p.t. These efficiencies were deemed acceptable for subsequent experiments. Then the GFP+ cells were cultured to validate the deletion.

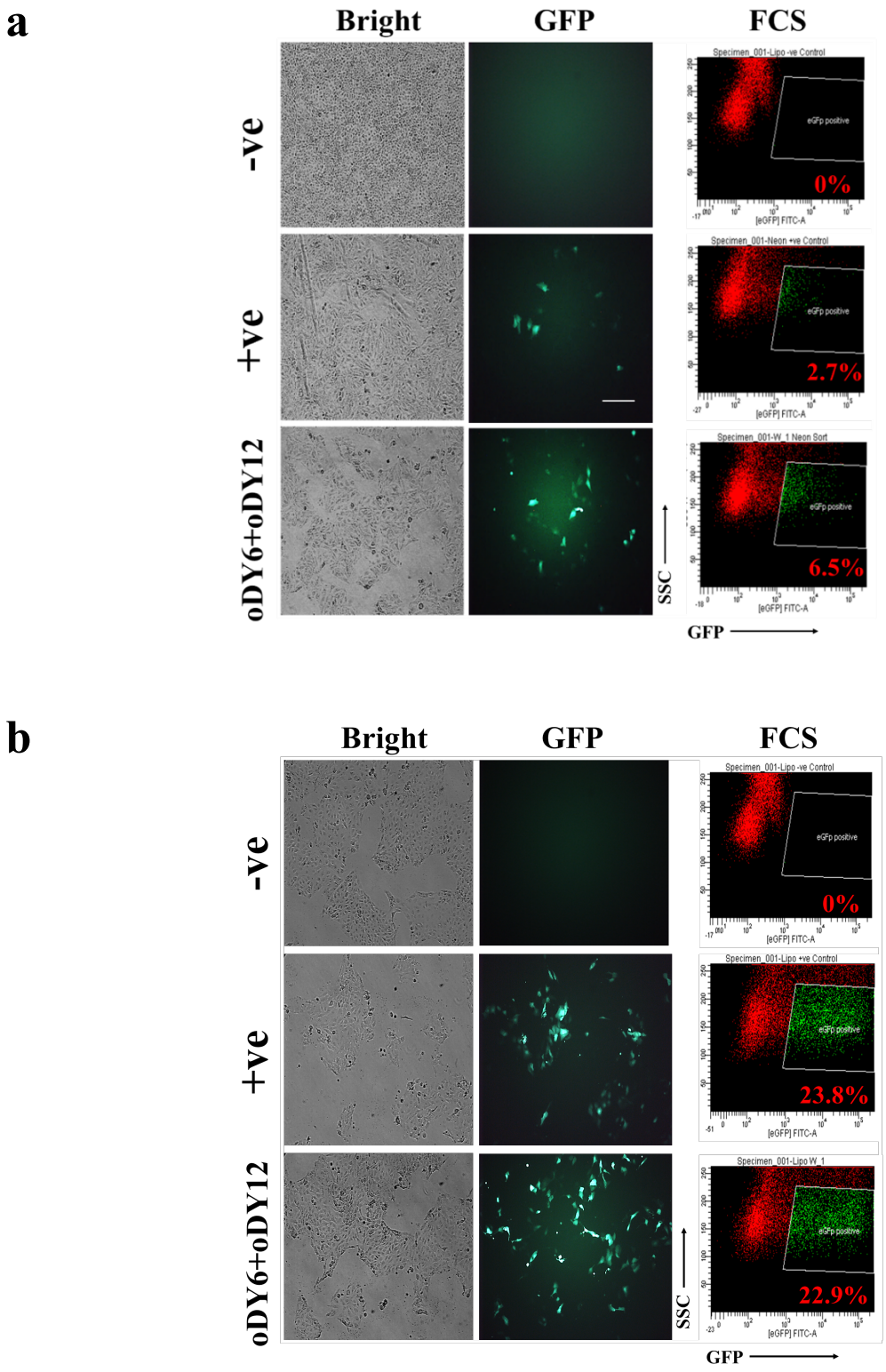


Figure 4. 12 Lipofectamine™ 2000 transfection optimization in NPTr cells. NPTr cells were analysed by GFP expression at 24 h.p.t in fluorescent microscopy (left), and FACS (right). Due to the gating was not proper, the actual transfection efficiency should be higher than the value shown on the image. Scale bar=100 µm.

4.3.3 Validation of homozygous deletion

After the transfection with Cas9 protein and sgRNA-containing plasmids (oDY6 and oDY12) (**Figure 4. 12**), and single cell isolation in 96-well plates, gDNAs were harvested from well-grown cell clones for PCR amplifications. The -/-ΔP candidates selected by PCRs were then processed by Sanger sequencing (**Section 2.3.1.8**) and Southern blot (**Section 2.3.1.11**).

GFP+ cells were individually sorted into one 96-well plate, followed by the cell clone expansion. gDNAs isolated from clones surviving expansion were used as templates for PCR. A pair of primers CoF1 and R7 spanning the target region was used to amplify a PCR product of 3.6 kb. R7, instead of CoR1, was used here because they were more efficient on amplifying PCR product more than 3 kb, comparing to CoF1 and CoR1 (Figure 4.10). Due to the absence of the ~2.9 kb target region, the PCR using these two primers was anticipated to yield a product ~0.7 kb, as shown in the mock gel **(Figure 4. 13-a, right panel)**. This PCR was termed as 'spanning PCR'. Among all the cell clones (**Figure 4. 13-b**), the negative control (-ve) showed no PCR products in the absence of DNA templates. The spanning PCR on the gDNA extracted from WT cells yielded an amplicon with a visualised size between 3.5 kb and 4 kb. The PCR product of clone 6 exhibited a WT allele (~3.6 kb). The PCR products of cell clones 1, 3, 4, 5, 8, 9 and 11 presented a mutant allele (~0.7 kb). They were thus prepared as candidates for further processing. We noted that the PCR on resulted in another faint band (~1.7 kb), suggesting that a mutation, for instance a random genomic integration occurred in the clone1 genome.

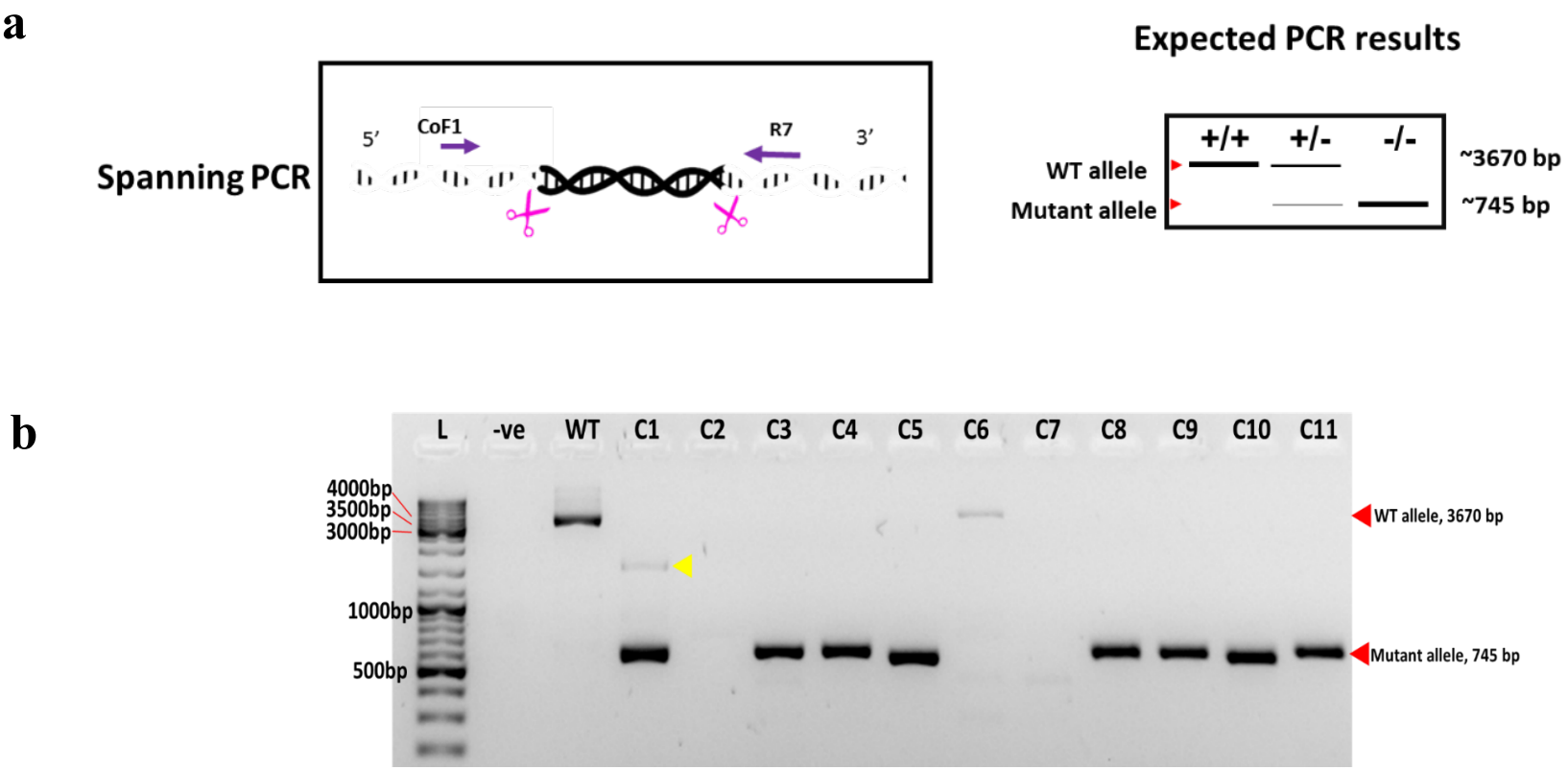


Figure 4. 13 Detection of 2.9 kb deletion by spanning PCR. (a) Schematic diagram showed that primers (CoF1 and R7) were used to amplify region spanning the target region. On the mock gel (right), a fragment of ~3.7 kb from a WT allele and/or ~0.7 kb from an edited allele were anticipated to be generated. CRISPR/Cas9s RNA guide sequences are depicted in two pink scissors. (b) The PCR results show clone1, 3, 4, 5, 7, 8, 10 and 11 harboured the editing. A PCR band in C1 lane was unexpected and indicated by a yellow arrow. It might be a random integration on the genome.

After that, the PCR band with the size of ~0.7 kb was cut out, purified and send for Sanger sequencing. Unexpectedly, successful sequencing results were only obtained from clone1, 3, and 4, and no trace files from other samples though repeated. The alignment is shown in **Figure4.14,** the sequencing of WT genome was used as a reference, and aligned with those of edited cells. The sequencing result showed the excision between the CRIPSR target sites, indicating the editing event was successful in these 1, 3, and 4 cell clones. It was noting that two alleles of clone 1 harboured different trace files, indicating that the bi-allelic event with each allele being unique occurred potentially.

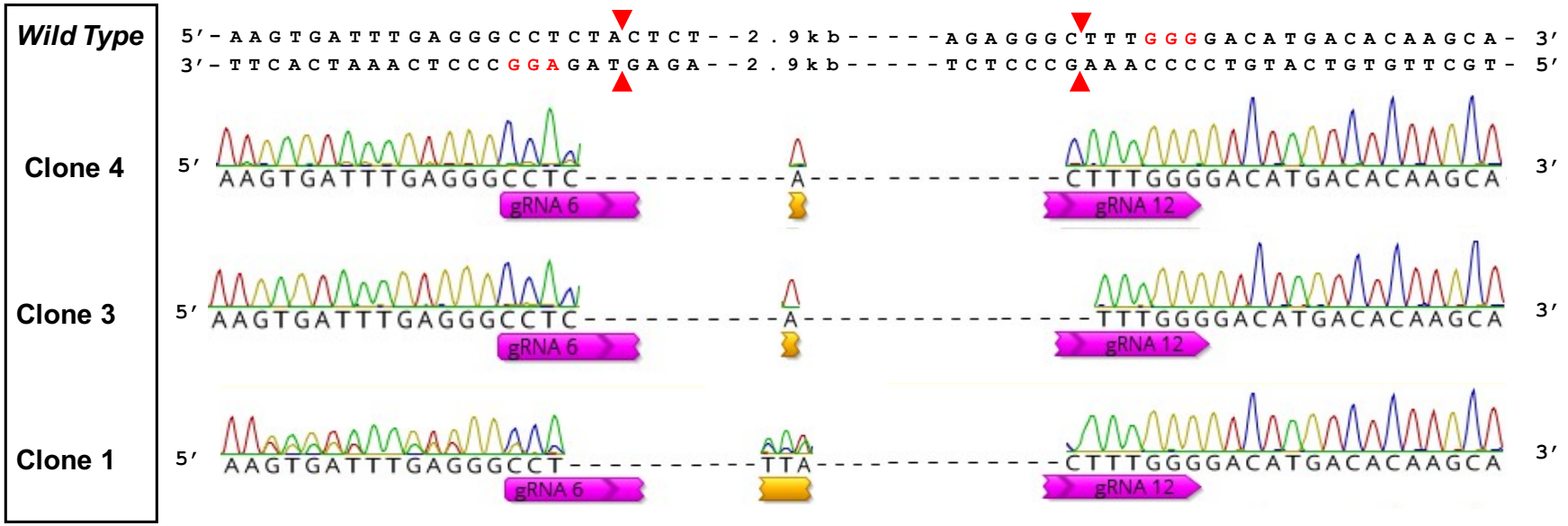


Figure 4. 14 Validation of editing by Sanger sequencing. The red nucleotides are PAM locus and the expected cutting site was pointed by the red arrows. Both clones present ~2.9 kb excision. Clone 1 harboured different sequencing results in two alleles, indicating an event of bi-allelic mutation or any other reasons which can be in further explored.

Considering the adequate performance of clones 4, 5, 8, 9, 10 and 11 in PCR results, we still included them in the Southern blot. The schematic overview of Southern blot is depicted in **Figure 4. 15a**, we used EcoR I enzyme to digest HWM gDNA isolated from each cell clones. An 11.3 kb fragment surrounding the target region was obtained. The homozygous edits were anticipated to have an ~8.4 kb allele in the presence of a ~2.9 kb excision. There was a mistake that the ladder of hybridization blot was not imaged properly. In order to measure the band sizes of the digested gDNAs on the blot, two blots had to be scaled horizontally according to the two undigested gDNA bands appeared in C11 sample (top of the blot, **Figure 4. 15-b**). The approximate sizes of the products on the hybridization blot could be reasonably speculated according to the ladder on the 0.7% agarose gel (**Figure 4. 15-b**, left panel). The result of hybridization blot (**Figure 4. 15-b**, right panel) showed Southern blot analysis failed in the C5, C8, C9, C10 and C11 samples (due to the poor enzyme digestion). There was only one fragment of ~11.3kb in WT, as a WT allele. A single fragment of ~8 kb appeared in C1, suggesting C1 was a homozygous allele (-/-). C3 and C4 appeared to have two bands, suggesting they were heterozygous edited (-/+).

These results suggest that C1 is a -/- cell clone and C3 is a -/+ cell clone. As several mistakes were made and discussed above, to achieve better outcomes of Southern blot, many improvements could be corrected accordingly. It was decided that for further experiments, more positive edited clones, an appropriate ladder, and more PCR amplifications were required.

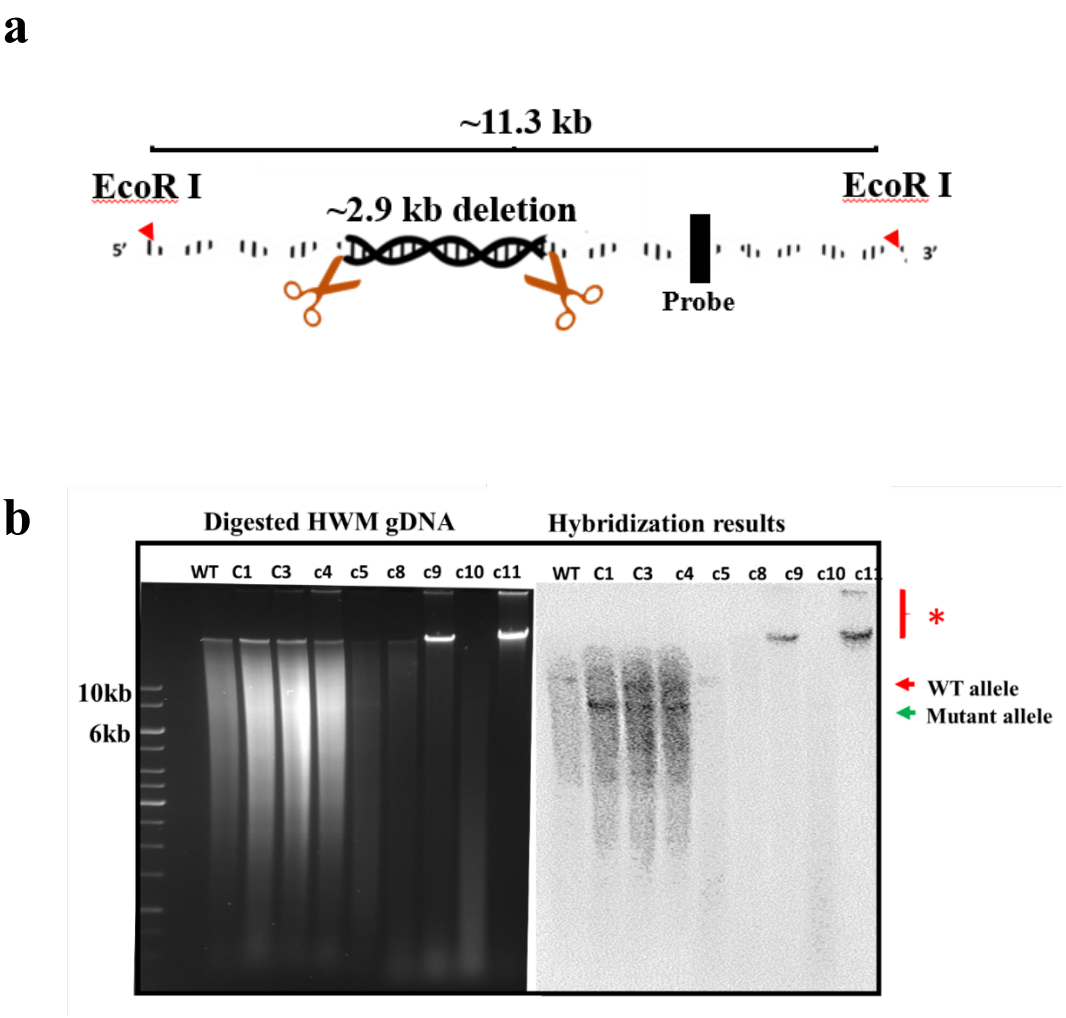


Figure 4. 15 Southern blot. (a) The schematic representation of the Southern blot. The probe (black box) 3' to the target locus and within an EcoR I-digested DNA fragment (~11.3 kb). The two sgRNAs are depicted by yellow scissors. A WT allele was ~11.3 kb and a mutant allele was ~8.4 kb. (b) Restriction digestion and hybridization results on candidate clones. EcoR I digested gDNA from each sample were hybridized with the probe. The hybridisation step was conducted by Claire Neil (Roslin Institute). 1 kb DNA ladder is on the left of the 0.7% agarose gel (left panel) but was excised prior to blotting. The blot was scaled horizontally by using the two undigested bands in C11 as a reference (red asterisk). Arrows indicated the expected band sizes for the WT and mutant alleles.

In the second round, GFP+ cells were individually sorted into 15 x 96-well plates, and the remaining GFP+ cells were enriched as a mixed pool (MP). We performed PCRs to detect the deletion in each clone. A schematic overview of the PCR screening strategy is shown in **Figure 4. 16**. Four rounds of PCR amplification were performed on the gDNA isolated from each clone surviving expansion, unless otherwise stated.

The first PCR reaction was performed spanning the target site with external primers designed to anneal ~300 bp upstream and downstream from the PAM sites of two sgRNAs (**a**, F1 and R7). The PCR product was anticipated to harbour ~3.5 kb in a WT allele, and ~0.6 kb amplicon in a mutant allele analysed on the 2% agarose gel. Two junction PCRs were performed using a pair of primers F2 and R2 or F9 and R9 (**b**). In the presence of a WT allele, fragments of ~0.9 kb and ~0.7 kb in size could be amplified, respectively. An internal PCR with primers F6 and R6 (**c**) was anticipated to produce a fragment of ~0.8 kb in the presence of a WT allele. Each mock gel is displayed on the right of each panels

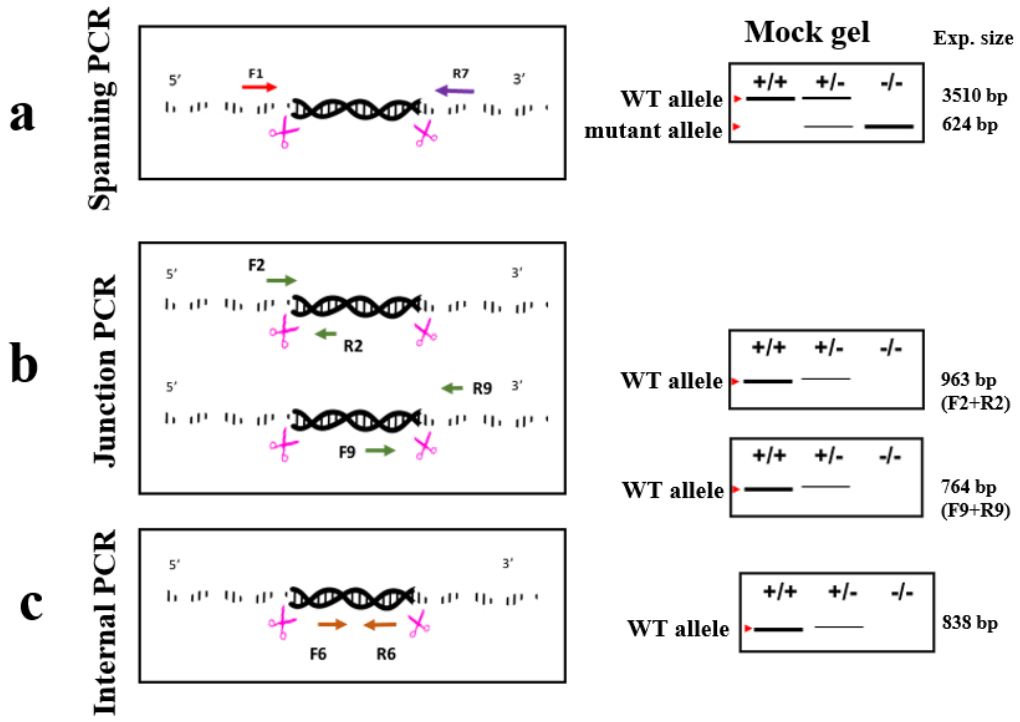


Figure 4. 16 PCR strategies for the validation of homozygous edits. (a)Spanning PCR is anticipated to generate a fragment of 3510 bp from a WT allele or 624 bp in an edited allele. (b)Junction PCRs are anticipated to generate a fragment of 963 bp (F2+R2) or 764 bp(F9+R9) in the presence of a WT allele. (c)Internal PCR would generate a fragment of 838 bp in the presence of a WT.
The half intensity of gel bands indicates heterogeneity.

In the result of spanning PCR (**Figure 4. 17**), the positive control (+ve), which was a paternal WT, and the MP sample failed to present a WT allele (~3.5 kb), indicating the PCR was suboptimal in amplifying the large WT fragment. Most of samples presented amplicons of 624 bp in size, indicating that most samples harboured a mutant allele.

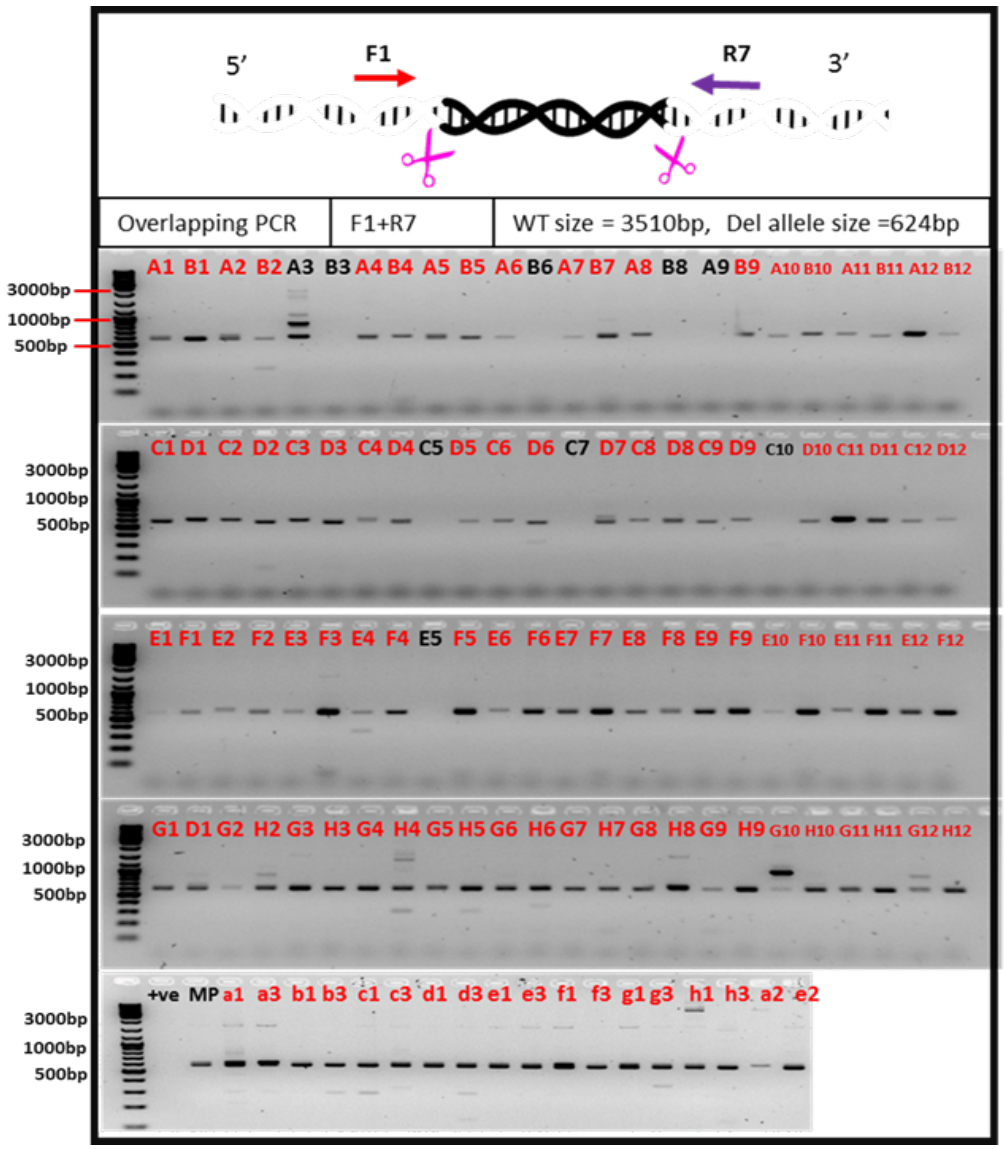


Figure 4. 17 Detection of deletion by spanning PCR. The samples were amplified using primers F1 and R7. The concentration of each extracted gDNA was normalized to approximately 20 ng/µl. The positive control (+ve) failed to show a WT allele, and the mixed pool (MP) showed only one mutant allele. Samples labelled in red harboured the successfully cleavage events.

Junction PCR was then performed using PCR primers flanking each CRISPR target site. The left junction PCR was using primers F2 and R2, and anticipated to result in a ~0.9 kb DNA product in the presence of a WT allele. As can be seen in **Figure 4. 18**, clones B3, B10, B12, C11, D12, F1, F4, F5, F7, F9, G2, G6, H7, H8, a1, a2, b1, b3, c1, c3, e1, f1, f3, g1, g3, h1, h3, e2 did not produce a PCR product. Some samples, such as F11 and H2, which presented faint bands, were not confirmed. This might be due to the leaky loading, or other reasons.

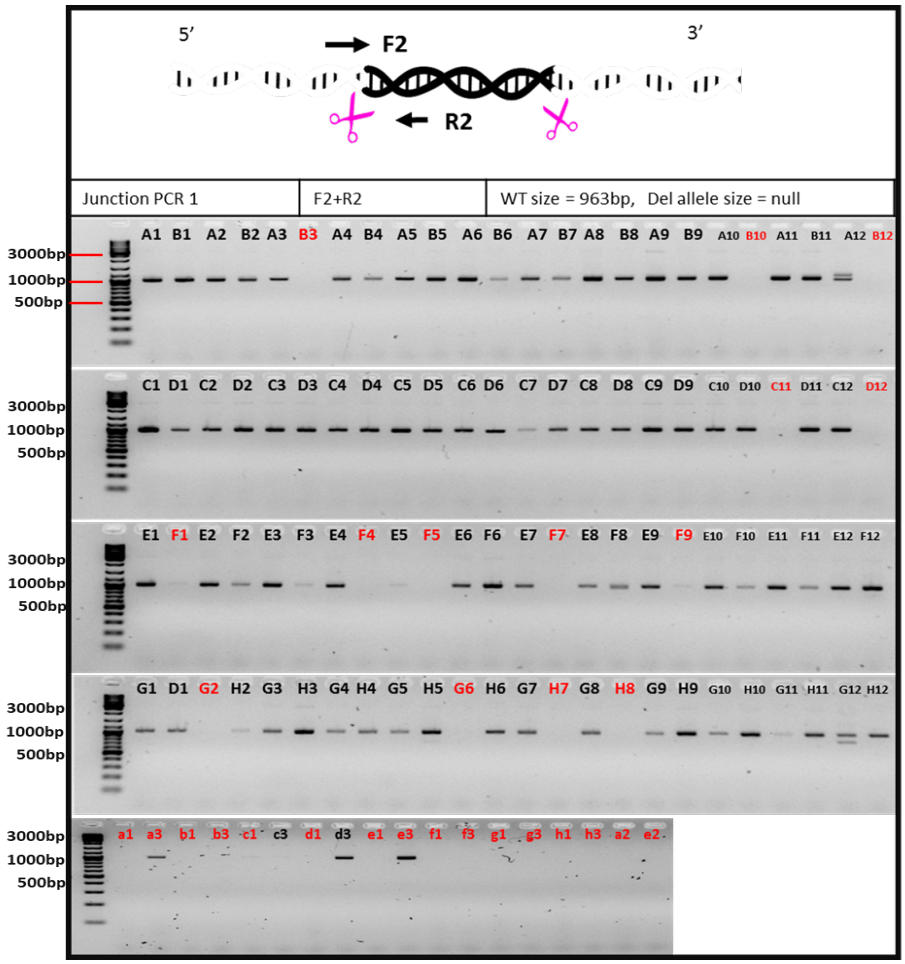


Figure 4. 18 Detection of deletion by left junction PCR. The samples were amplified using primers F2 and R2. WT allele was anticipated to produce a product which size was ~0.9 kb, and -/- edits would not produce any products. Samples identified as -/- candidates were labelled in red. WT and MP samples were forgotten to be included in the junction PCR inadvertently.

In the right junction PCR (**c**), clones A2, F1, G2, E3, B3, E5, F4, F5, H6, C7, F7, H7, F9, D9, A10, B10, C11, F12, D12, a1, a2, a3, b1, d1, e1, f1, f3, g1, g3, h1, h3 did not produce a PCR product, in the presence of a WT allele. Other samples, such as H4, E11, and G11, which showed faint bands, were not confirmed. This might be due to the leaky loading, or other reasons.

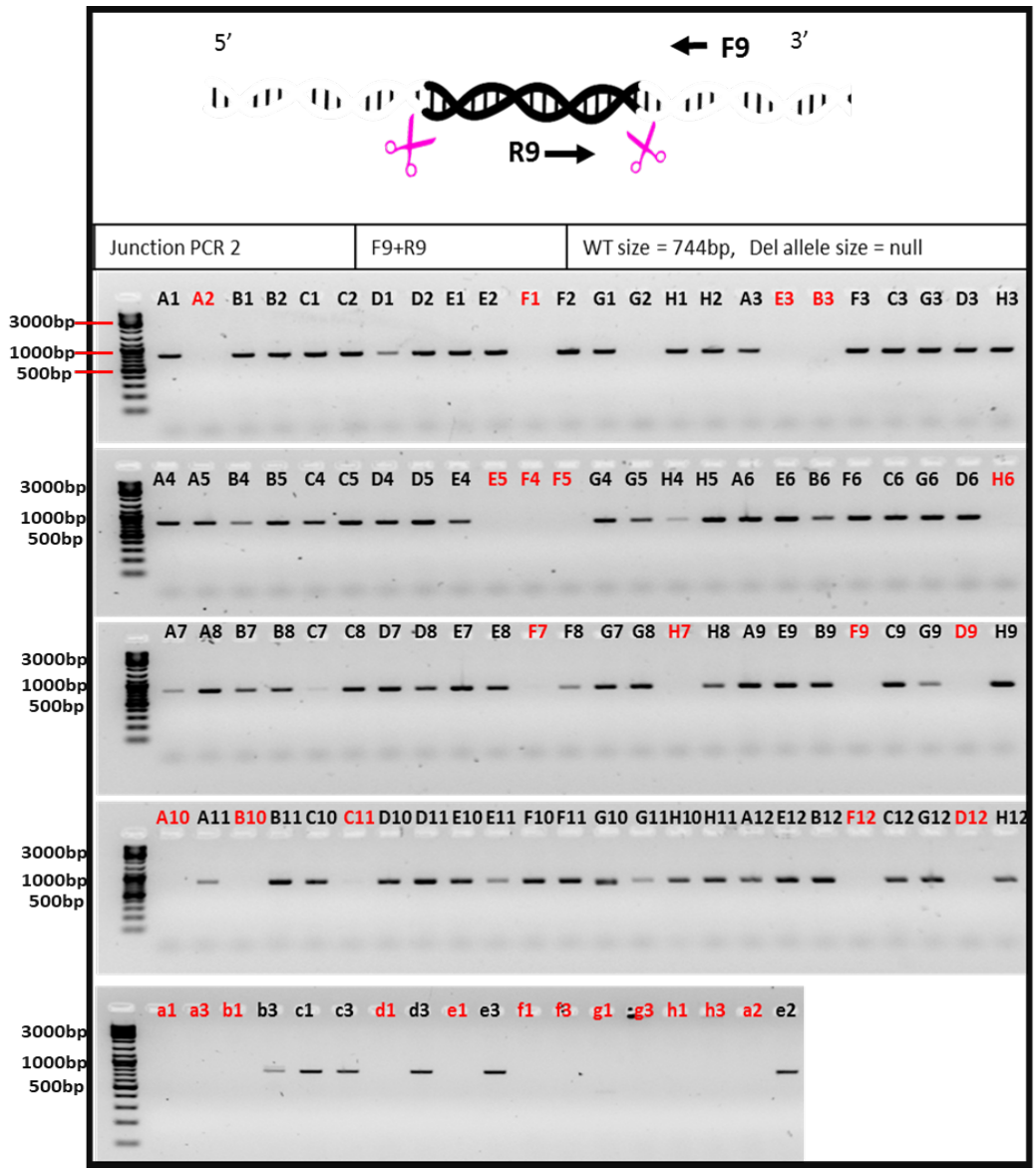


Figure 4. 19 Detection of deletion by right junction PCR. The samples were amplified using primers F9 and R9. WT allele was anticipated to produce a product which size was ~0.7 kb, and -/- edits would not produce any products. All the eligible samples were red labelled. WT and MP samples were forgotten to be included in the junction PCR inadvertently.

The internal PCR was then performed using primers within the target region. It was anticipated a fragment of ~0.8 kb in size would be generated in the presence of a WT allele, but that would not occur in homozygous clones. **Figure 4. 20** shows that clones B3, B10, C11, D12, F1, E2, F4, F5, F7, F9, G2, H7, a1, f3 and e2 produced either no PCR products or a very faint signal.

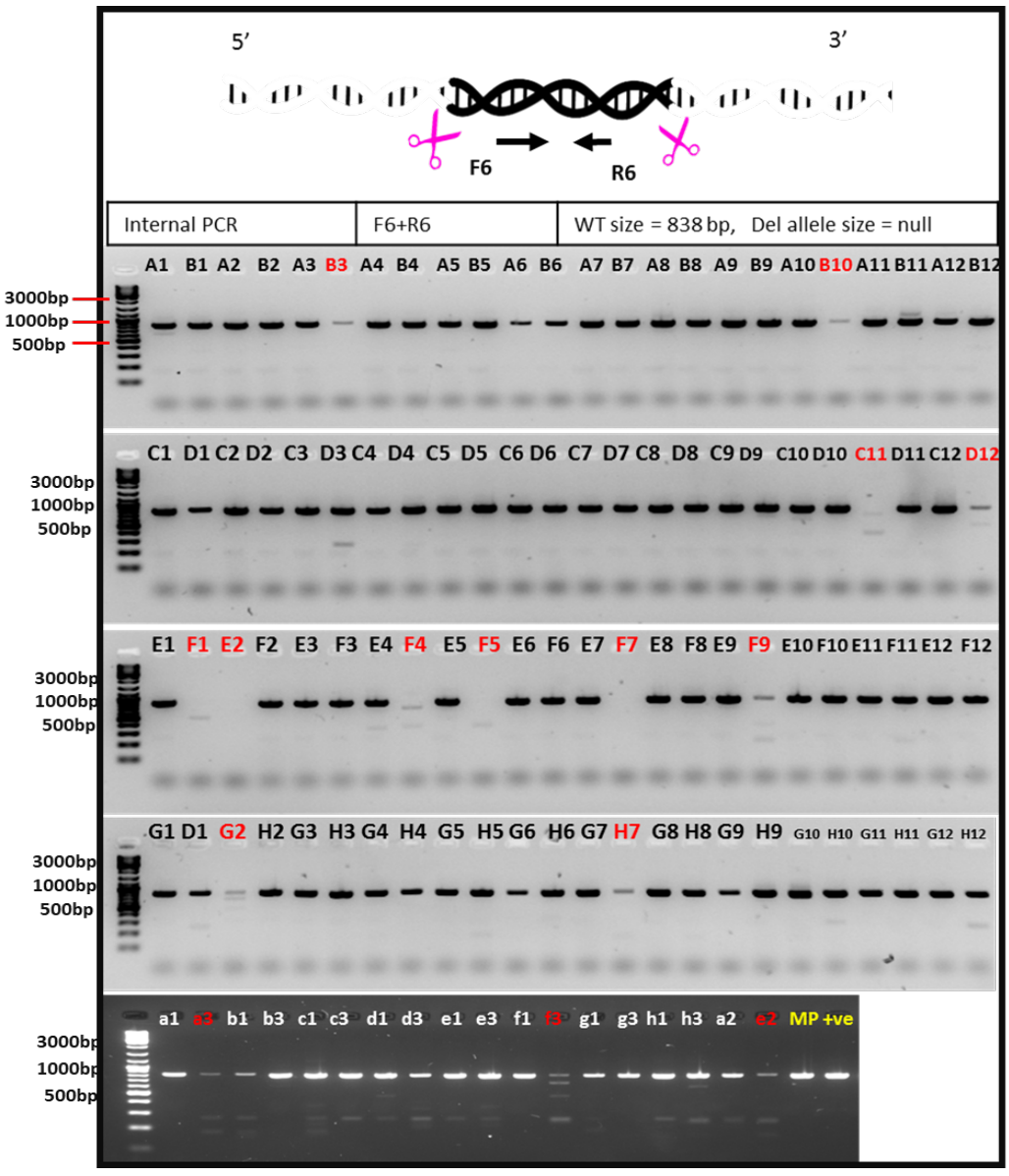


Figure 4. 20 Detection of deletion by internal PCR. The samples were amplified using primers F6 and R6. WT allele was anticipated to produce a product which size was ~0.7 kb. -/- edits would not produce any products. Most of the PCR samples show a ~0.8 kb amplicon. All the eligible samples were red labelled.

After the PCRs on all clones, B10, C11, F1 and F7 were eligible to four rounds of PCR screening. We also involved B3, D12, E2, F4, F5, F9, G2, H7, a1, a2, a3, e1 and f3 as -/- candidates as they were eligible to all PCRs except the internal PCR, and the reason would be discussed in **Section 4. 5**. In fact, B3 actually included inadvertently as B3 did not present a mutant allele in the spanning PCR.

All PCR products were sent for Sanger sequencing analysis (**Section 2.3.1.8**). Selectively, the sequencing results of -/- candidates are shown **below**. Apart from B3, D10, a1, e1 and f3, other clones exhibited successful excision of DNA sequencing. The sequencing data and trace files of all clones for sequencing were shown in **Appendix 10**. We observed the contig aligned with the excised nucleotides, indicating they are homozygous deletion.

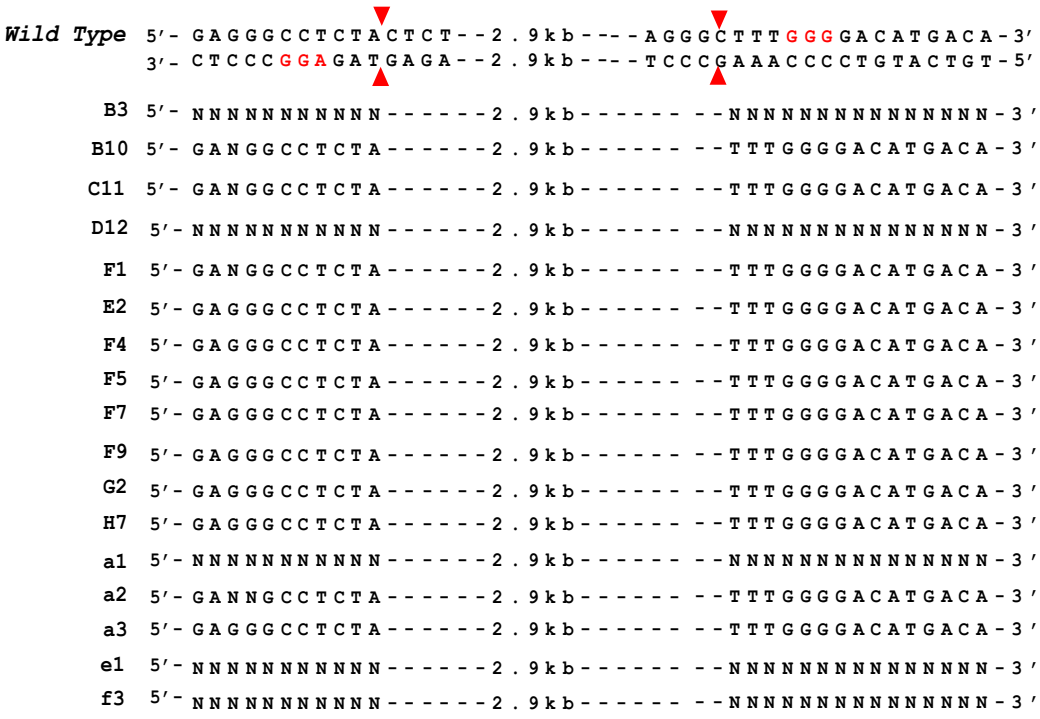


Figure 4. 21 Validation of editing by Sanger sequencing. The samples selected here were identified to be homozygous by PCR assays. Compared with WT sequence (top line), samples (except B3, D12, a1, e1, and f3) show the correct excision in the genome. All trace files are in the **Appendix 10**.

Southern blot analysis was then performed as described previously (**Section 2.3.1.11**). As shown in **Figure 4. 22**, the EcoR I digested DNA appeared as a smear on the 0.7% agarose gel (**a**). In the result of hybridisation (**b**), WT was anticipated to be a ~11.3 kb; and -/- edits presented as a ~ 8.4 kb product. Among all the candidates, four clones (D12, F4, F7 and a3) were identified to be homozygous edited and one clone (f3) were identified to be heterozygous edited. B3, B10, C11, E2, F5, F10, H7, a2 were identified to be WT, which results were not consistent with previous PCRs and Sanger sequencing results. Moreover, F1 contained an extra band, which was thought to be as ~20 kb according to the ladder, suggesting a random integration in the genome.

Taken together, C1, which was identified in the first round of selection, and D12, F4, F7, and a3, which were identified in the second round of selection, were confirmed to be homozygous edited cells. There were cultured for the phenotype assessment and the IAV infection (**Chapter 5**).

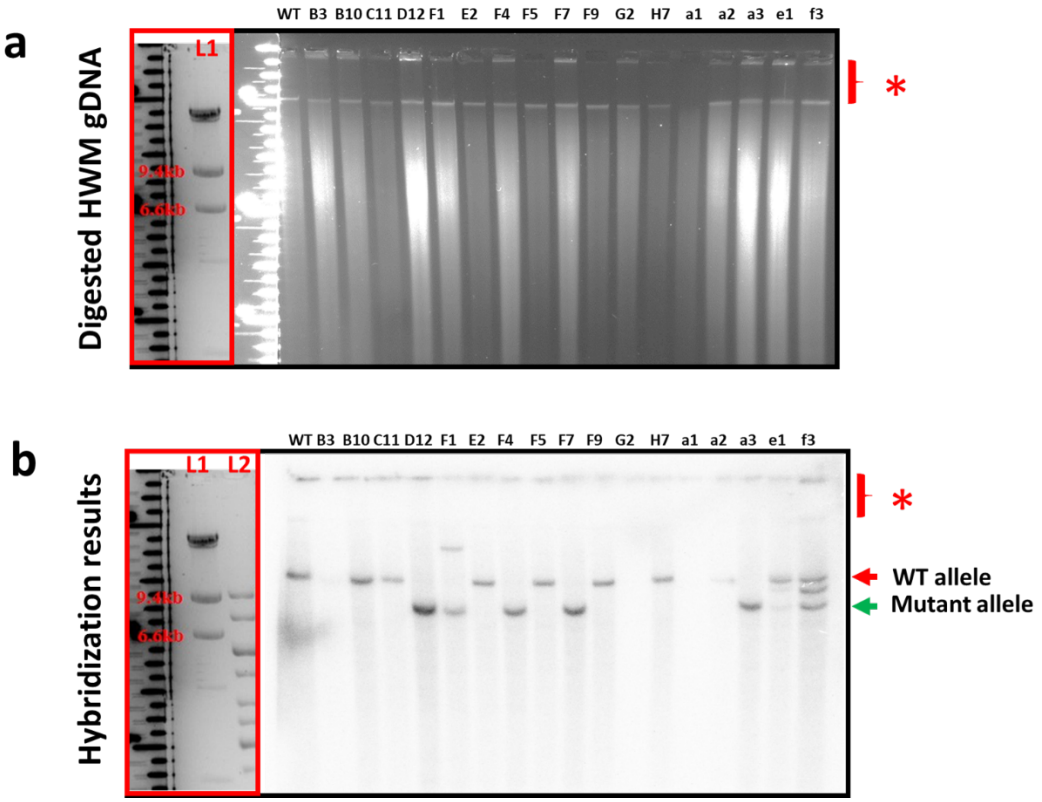


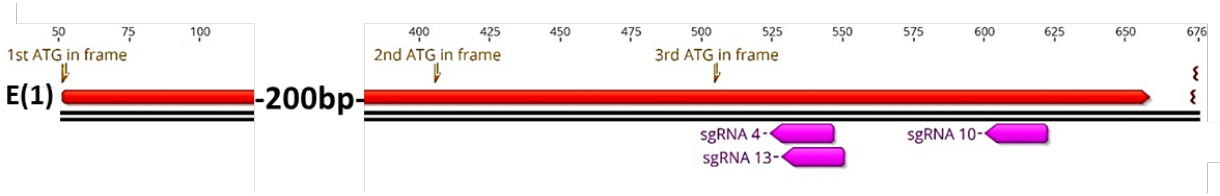
Figure 4. 22 Southern blot. (a) EcoR I digested HMW gDNA appeared to be a smear on a 0.7% agarose gel, followed by the transferring and blocking steps. (b) The blot was hybridised with a radiolabelled PCR product. From three rounds of PCRs and sequencing selection based, four -/- edited clones (D12, F4, F7 and a3) and one -/+ edited clone (f3) was identified. The gel and the blot were scaled horizontally according to the well and the biggest size of DNA (red asterisk). L1: Lambda DNA/Hind III marker. L2: 1 kb DNA ladder.

4.4 Result 2—functional knockout of ST6GAL1

4.4.1 CRISPR reagent construction

sgRNA was designed according to the strategy we described in **Section 4.2**. Each annealed oligos (sgRNA) were subcloned into the Cas9-contained pSL66 vector. Sanger sequencing was used to confirm the inserted sequence in each CRISPR plasmid (**Figure 4. 23-b**). In the figure, the upper line (pSL_U6-oSL35) was the WT sequence with a block of 20N denoting the intended position of the inserted oligonucleotide duplex. All colonies picked represented successful CRISPR assembly.

a



b

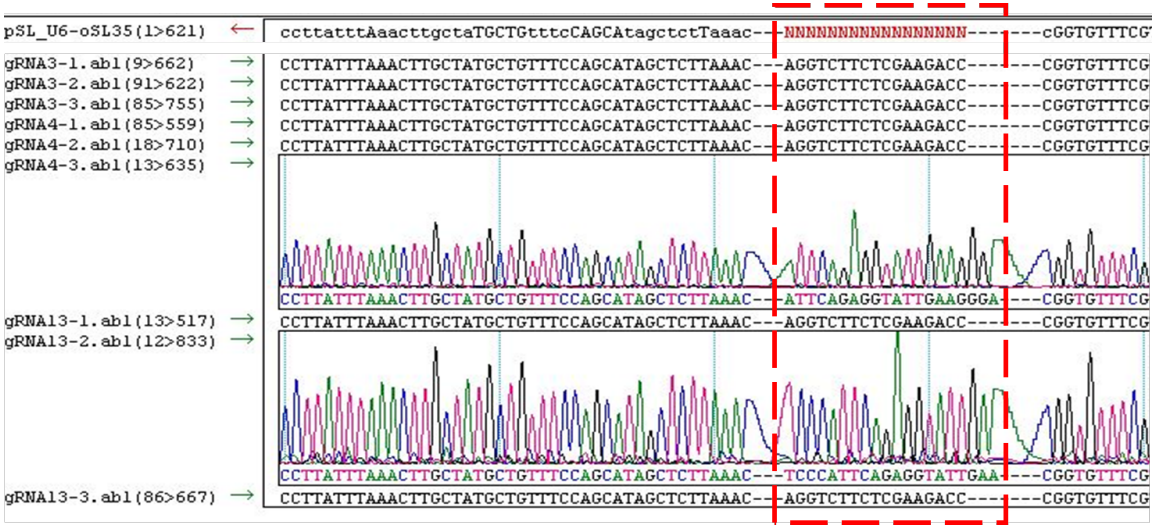


Figure 4. 23 CRISPR reagent assembly. (a) The schematic diagram depicted sgRNA4, 10, and 13, locating downstream of the third ATG which is in the same open reading frame with the first ATG in coding exon1. (b) Sanger sequencing result showed that sgRNA was successfully ligated into pSL66 vector, and it is corresponding with 20 'N' in reference sequence.

4.4.2 Transfection and CRISPR guides validation

Following the plasmid maxi-prep, 1.5 µg CRISPR/Cas9 reagents containing sgRNA4, 10 and 13, assigned as P4, P10, and P13, was respectively transfected in NPTr cells using the same protocol as described in **Section 4.3.2.4.** GFP+ cells were enriched for the assessment of transfection efficiency. The transfection results showed there were no GFP fluorescent signals in the mock transfected control (-ve). The positive control (+ve) with 1.5 µg pSL66 plasmid was without any sgRNA insertions.

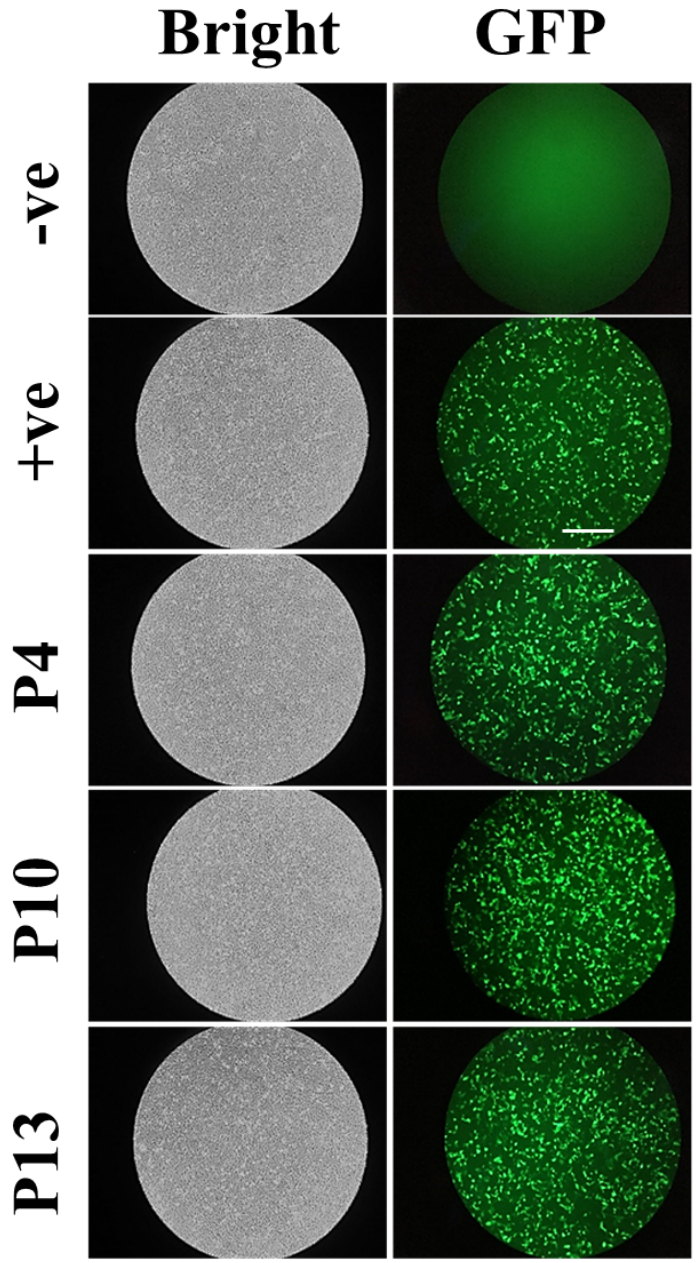


Figure 4. 24 Analysis of transfection in NPTr cells. (a) Bright and fluorescent images of cells transfected with CRISPR plasmids. Because of the computer issues during the experiment, the images were taken using a mobile phone through the ocular lens. Scale bar=100 µm.

4.4.2.1 Mismatch cleavage assay by T7E1 and TIDE assay

T7 Endonuclease 1 (T7E1) and TIDE assay were performed to test the cutting efficiency of sgRNA4, 10, and 13. The T7E1 assay was performed as described

previously (**Section 2.3.1.10.3**). Before that, a fragment of 626 bp was amplified using gDNA isolated from each GFP+ population using primers E1F1 and E1R1. Each PCR products were mocked-digested and T7E1-digested, respectively. All products were visualized on a 1% agarose gel (**Figure 4. 25-b**). The heteroduplex DNA was recognised and cleaved by T7 endonuclease I. We observed no cutting activity of sgRNA13.

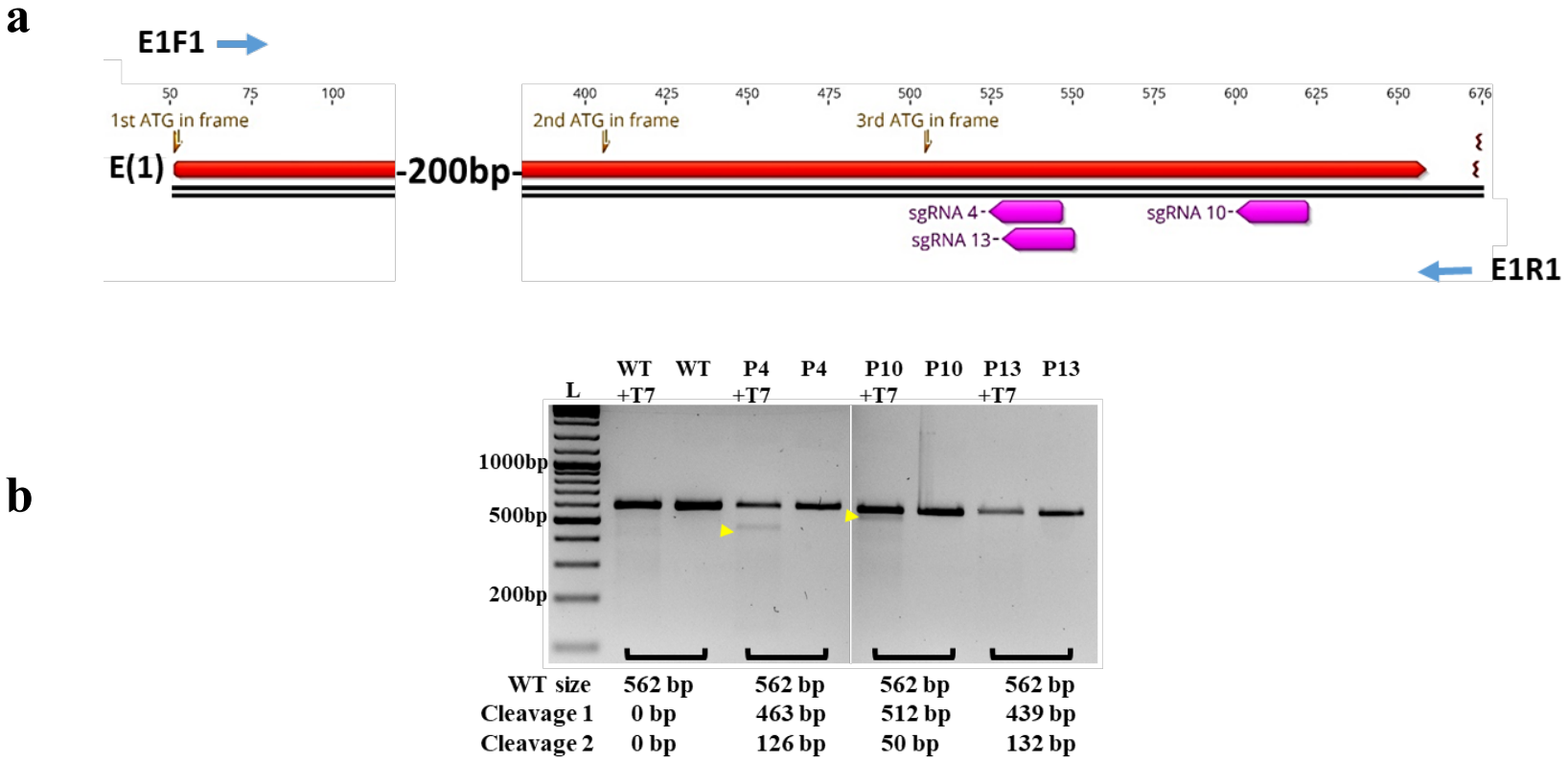


Figure 4. 25 T7E1 assay. (a) The schematic overview of sgRNAs and the primer pair E1F1 and E1R1. (b) Each amplified PCR product was incubated with either T7EI or water (control). The table shows the expected sizes of cleaved products, which are indicated by yellow arrows on the gel.

The TIDE assay was then applied to estimate the editing efficiency of sgRNA4. This is based on the decomposition of the sequence trace to estimate indels introduced by NHEJ during DSBR (Brinkman et al., 2014). As seen in **Figure 4. 26**, the result of TIDE assay showed sgRNA4 had 10.2% cutting efficiency.The predicted in/dels were shown in upper panel. In the middle panel, the region of decomposition was depicted as a disruption of the tracing. The deviating sequence downstream the break site of sgRNA4 was aberrant with the reference sequence of WT (vertical dotted line).The bottom panel exhibited the contigs in P4 aligned with the WT sequence

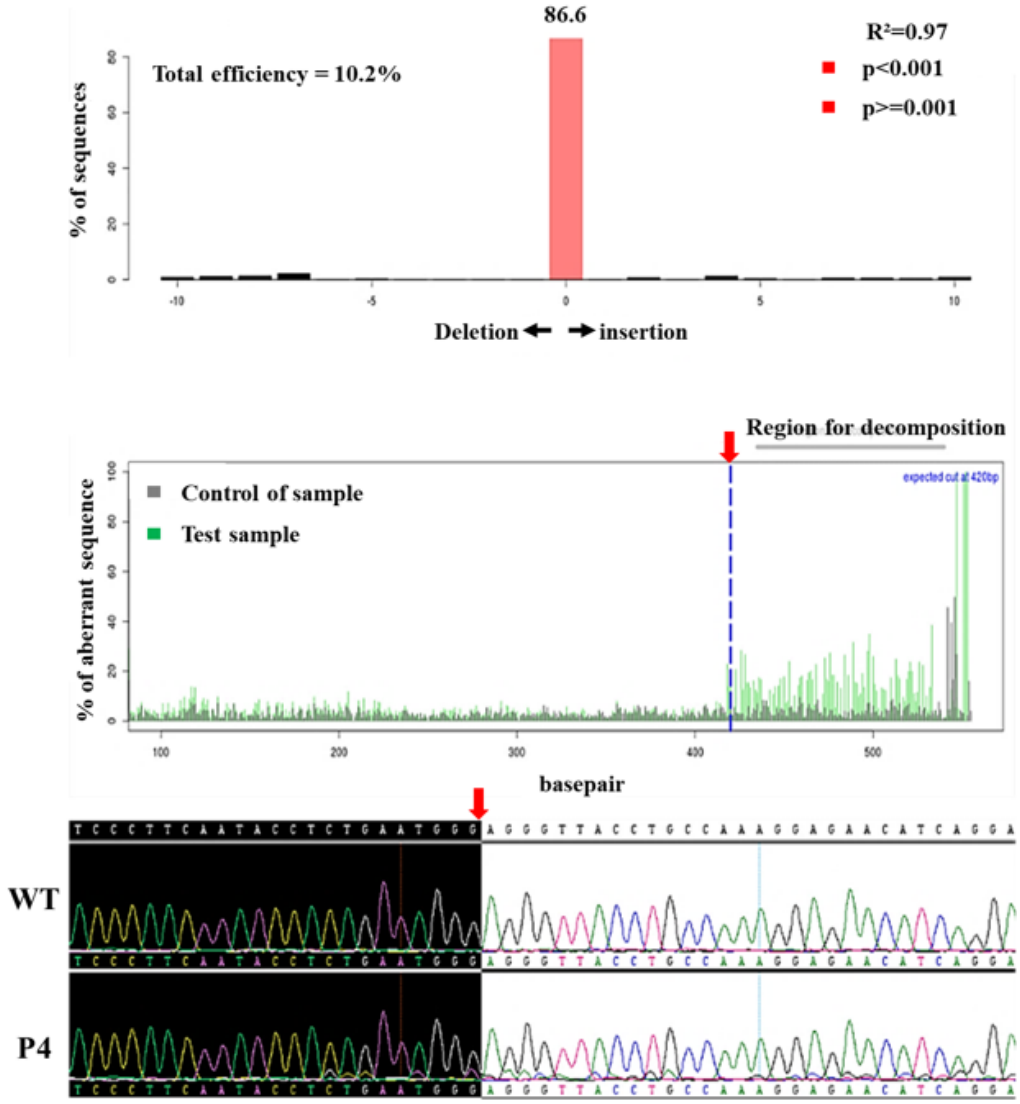


Figure 4. 26 TIDE assay. The panel presents the comparison of the tracing sequence between the reference (WT, top trace) and P4 (Mutant, bottom trace). The aberrant sequence signal and in/dels spectrum (R2) are shown in the histogram. The cutting efficiency was calculated as percentages automatically by the software.

4.4.2.2 The design of the ssODN template

As sgRNA4 was identified to be the best guide by both T7E1 and TIDE assay, a single strand donor DNA template was designed to introduce a homology directed repair at the cut site of sgRNA4. As shown below, a single-stranded donor oligonucleotides (ssODN) donor template (130 nt) sequence with 40 nt of 5' homology to the DSB and 90 nt of 3' homology. This asymmetrically design has previously been reported to result in a high HDR efficiency (Richardson et al., 2016). As shown in **Figure 4. 27**, a mutation was introduced to convert the PAM sequence GGG to TCT (GGG>TCT). This mutation was intended to have an effect of introducing a novel EcoRI restriction site (red arrow), destroying the endogenous PAM such that re-cutting following successful HDR is unlikely (Paquet et al., 2016), and introducing a premature TAG stop codon in frame (black box with *).

Figure 4. 27 The HDR template. The lower oligo strand is the HDR template. Grey shadow is stop codon and lined sequence is the restiction enzyme site (EcoR I). Red shadow is the PAM mutation (GGG>TCT).

4.4.3 Validation of editing

4.4.3.1 Co-transfection of the CRISPR reagent and ssODN template

1.5 µg P4 or 1.5 µg P4 plus 1.5 µg HDR template were chemically transfected into NPTr cells with optimised transfection conditions (**Section 4.3.2.4**). The fluorescent images at 24 h.p.t are shown in **Figure 4. 28**. The transfected cell population with both P4 and the HDR template exhibited fewer GFP+ cells than those transfected with only P4.

To assess the HDR efficiency, the cells transfected with sgRNA4 and HDR were cultured, and gDNA isolated, followed by the incubation with EcoR I and T7EI, separately. No EcoR I digestion products were observed (**Figure 4. 28-b**), indicating that if HDR did occur it was at very low levels. The T7E1-digested PCR products did show the anticipated cleave profile, corresponding to the previous result (**Figure 4. 25-b**). Taken together, this suggests that NHEJ is the primary method of DSBR following these transfections, and that producing a frameshift mutation via NHEJ may be a more viable approach than the introduction of a precise missense mutation by HDR. Therefore, GFP+ transfected cells with sgRNA4 were individually seeded into 9 x 96-well plates by FACS for cell expansion.

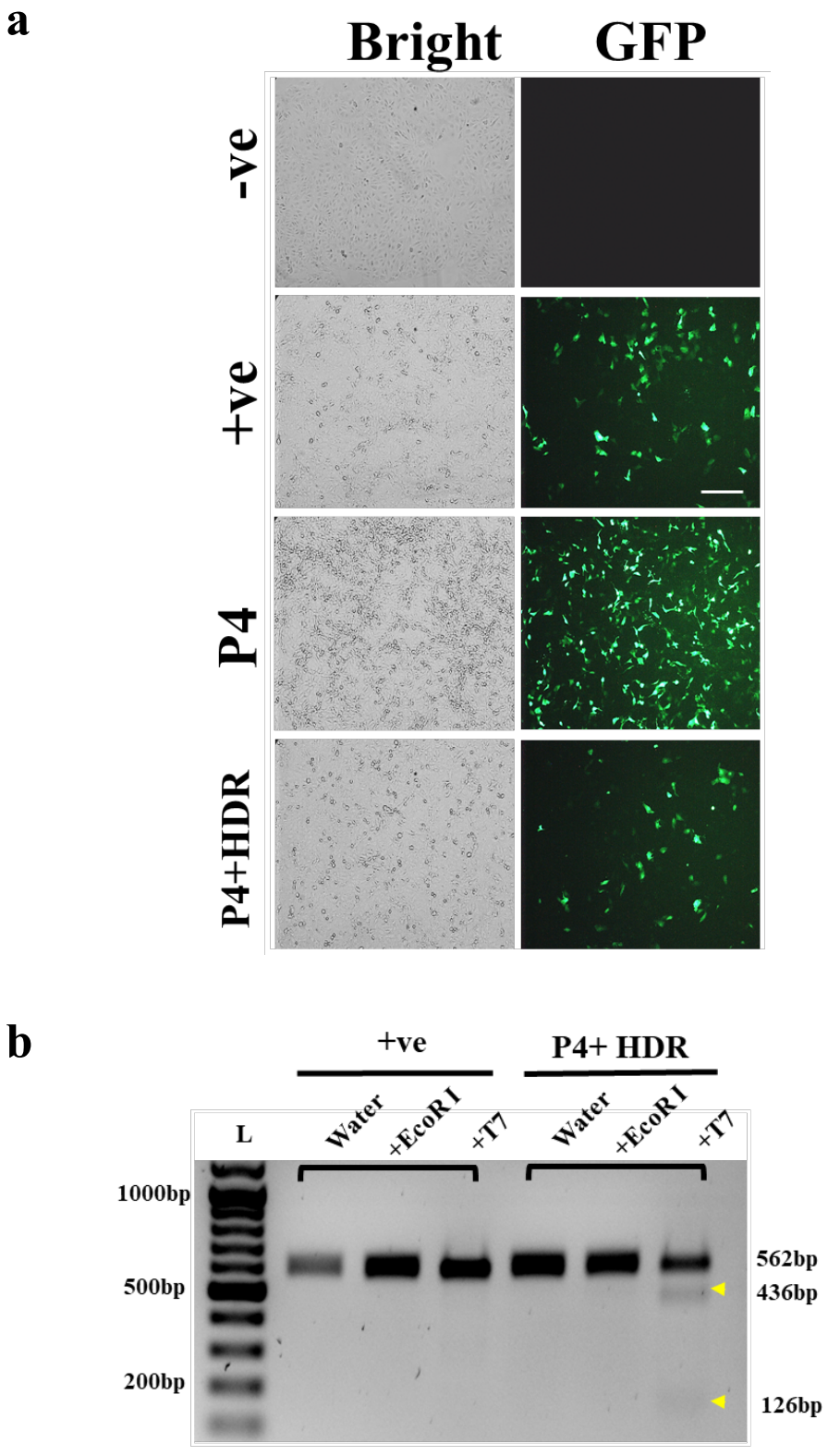


Figure 4. 2 Co-transfection and HDR validation. (a) Cells were transfected with P4 with/without the HDR template, and visualized at 24 h.p.t. GFP+ cells were single cell sorted into 9 x 96 well plates for expansion. (b) gDNAs of WT and P4+HDR transfected cells were amplified using the primer shown in **Figure 4.25-a**, followed by adding water, EcoR I and T7E1.The cleaved products were pointed by yellow asterisk.

4.4.3.2 Validation of editing in ΔCD cells

4.4.3.2.1 PCR-based assay

After approximately 50 days' clonal expansion, approximately 120 cell clones were confluent. gDNA was isolated from each clonal candidate using the QuickExtract™ approach (**Section 2.3.1.2.2**) for PCR amplification using primers E1F1 and E1R1, resulting in an amplicon with an anticipated size of 562 bp from WT cells, as shown in **Figure 4. 29**. The water control (-ve) showed no amplicons. All transfected samples labelled in red presented the connect size of band. Consequently, all the successful amplified PCR products were sent for Sanger sequencing.

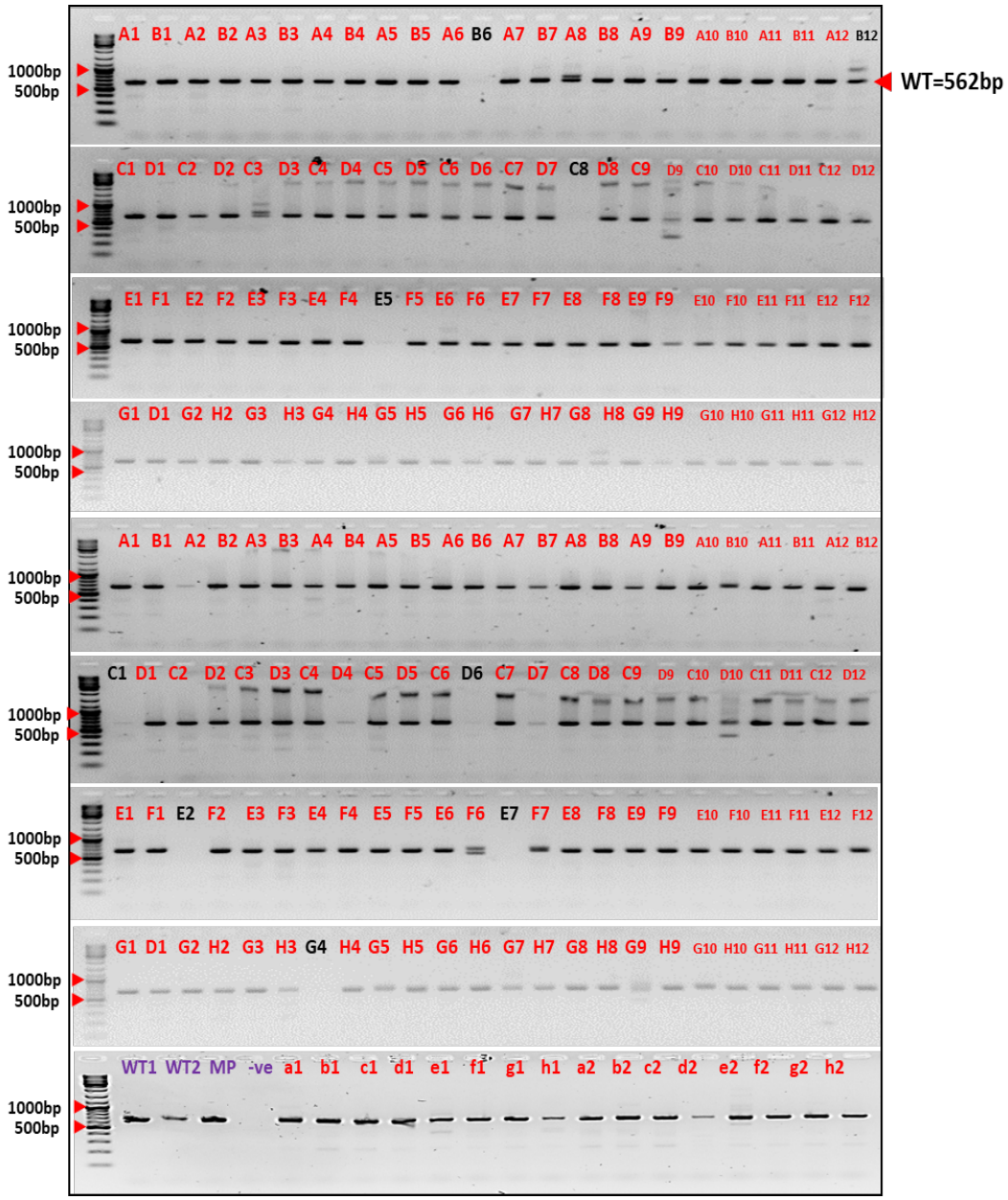


Figure 4. 29 Spanning PCR. gDNA was extracted from each clone using QuickExtract™ reagent. ~20 ng gDNA was used for PCR amplification using primers E1F1 and E1R1, resulting in a 562 bp product. A negative control (-ve), and three positive controls (two WT and one mixed pool) were included.

4.4.3.2.2 Sanger sequencing on ΔCD clones

The alignment of sequencing surrounding the target site between reference (WT) and homozygous edits is shown below. Raw sequencing data for all clones are in **Appendix 11**. Below each DNA sequence is the corresponding protein sequence. The event of the premature termination (the stop codon) is highlighted with an asterisk. It was noting that clone H10 and A1, the PAM sequence (GGG) was changed to TCT (**Figure 4. 30** and **Appendix 12**). A premature stop codon (TAG) was thus introduced, leading to a termination of translation successfully as described previously (**Section 4.4.2.2**). Therefore, we could conclude that the HDR event in H10 and A1 had a successful HDR event, though we did not observe the anticipated ECoR I digested product on the 1% agarose gel (**Figure 4. 28-b**). In cell clone F6 and G2, we detected a deletion of 55 bp and 8 bp, respectively. In other cell clones, they all presented 1 or 2 bp insertions within the PAM, resulting in a frameshift mutation.

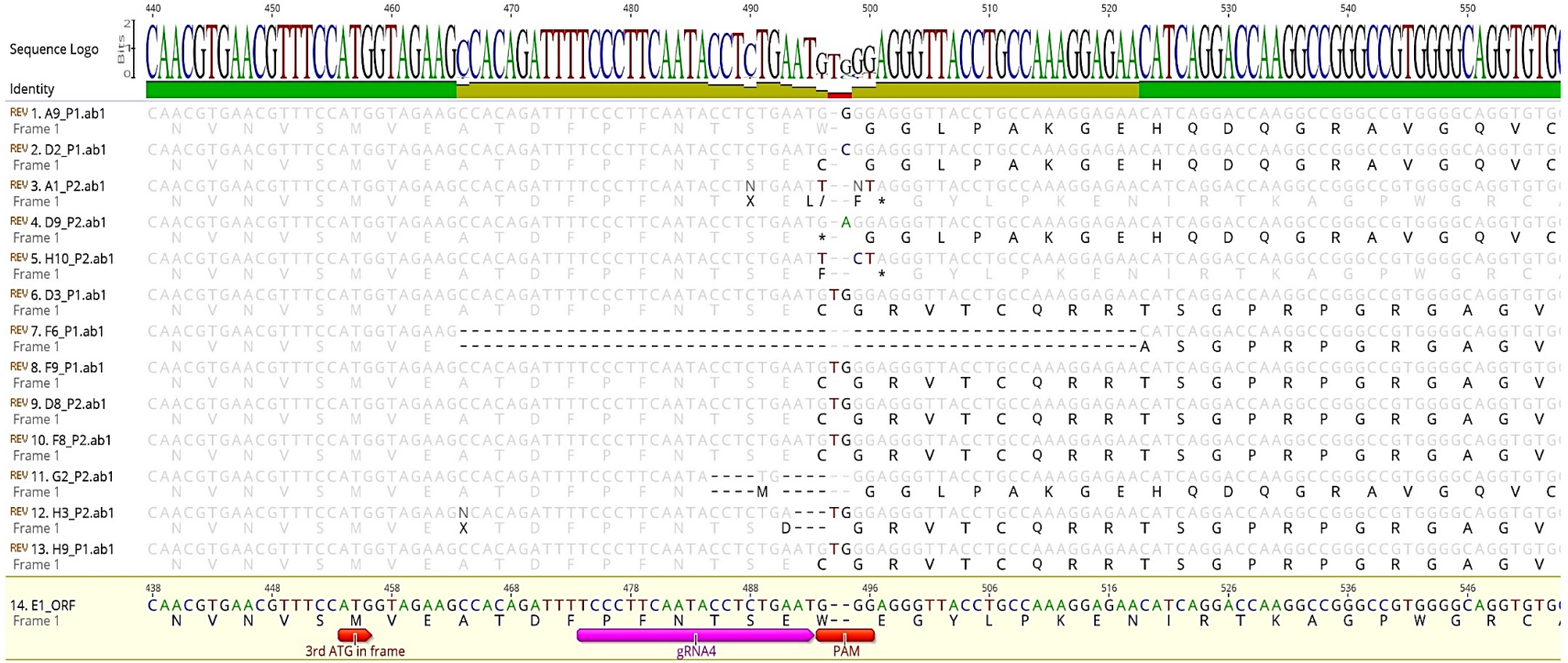


Figure 4. 30 Sanger sequencing on ΔCD clones. The reference sequence is at the bottom (E1_ORF). The DNA identity is shown in grey. sgRNA4 and PAM sequence (GGG) are labelled in pink and red, respectively. * represents a premature stop codon; ----- represents the deletion of sequence. Trace files of these clones are shown in **Appendix 12.**

4.5 Discussion

In this chapter, we evaluated two CRISPR/Cas9-based strategies to precisely edit the ST6GAL1 gene. We then identified homozygous edited clones after gene editing experiments. In the first strategy, we generated a homozygous NPTr cell line, -/-ΔP, which harboured an altered transcriptional profiling of ST6GAL1 in airway cells. In the second strategy, we created functional knockout of ST6GAL1 in NPTr cells, -/-ΔCD, by introducing a frameshift mutation in the coding sequence of ST6GAL1.

We now discuss how we optimised cell transfection efficiency. In particular, we preliminarily applied the Neon^{TM} electroporation system to transfect sgRNA/CRISPR plasmids in NPTr cells with the optimised (**Section 4.3.2.1**). However, the transfection efficiency for NPTr cells turned out to be extremely low with these conditions using electroporation. Therefore, we used the Lipofectamine® 2000 transfection system instead, which has been widely applied for the delivery of CRISPR/Cas9 constructs into mammalian cells (Yu et al., 2016). A much higher transfection efficiency was achieved (**Figure 4.12-b**). Alternatively, we could use the viral delivery to send Cas9/sgRNA components into NPTr cells, such as lentivirus. We could transduce LV vectors, among which sgRNA and Cas9 were integrated in one plasmid (Sanjana et al., 2014). However, for this to work properly, both the titres of lentivirus and the size of lentiviral vectors have to be optimised and considereed, which can be time consuming, labour intensive and cost ineffective. Therefore, Lipofectamine® 2000 transfection system was still the preferred option in this experiment.

All primers and annealing temperatures should be optimised prior to T7E1 assay. For example, if the primers and the annealing temperatures in PCR amplifications are suboptimal, then the cleavage products of T7E1 assay would not be discriminated from other non-specific PCR products or SNP events (**Figure 4. 8-b**). Therefore, PCR conditions should be optimised.

Validating the modification to the ST6GAL1 gene in -/-ΔP cells is crucial. To this end, we used four rounds of PCR amplifications, Sanger sequencing, and Southern blot. Spanning PCR was the first one applied to verify whether the deletion occurred. 106 out of 114 clones showed a band of the expected size of the mutant allele (see **Figure 4. 17**), indicating that these clones were homozygous or heterozygous mutations. Sanger sequencing analysis was consistent with the spanning PCR result (see **Appendix 10**). However, the products of the internal PCR (**Figure 4. 21**) which amplified a region within the ~2.9 kb deletion segment could still be observed in most of the clones, indicating that they were heterozygous mutations. Such a clone was excluded. Unexpectedly, B10 and C11 which were identified as -/-ΔP in all four PCR amplifications appeared to be WT type in the Southern blot analysis. As Southern blot is more sensitive and more accurate in ensuring the genotyping, we excluded in B10 and C11 for the phenotyping assessment. Furthermore, it was worth noting that there was one extra band appearing in F1, which may be due to a random integration in the genome. This is because CRISPR/Cas9 system can cause DNA rearrangements surrounding the target locus. Sometimes an unrelated DNA segment could be incorporated into the chromosome (Kosicki et al., 2018).

For the second approach, it is worth discussing why we designed the sgRNA downstream the third ATG codon instead of the first one, which is in the same open

reading frame with third one. The rationale behind this is that the third coding start (GA**G**C<u>ATG</u>**A**) has a strong 'Kozak sequence', while the first one (CA**T**T<u>ATG</u>**A**) (Gene ID: 100302026, NCBI), is empirically predicted to be a weak Kozak. The 'Kozak consensus sequence' is associated with the efficiency of translation (Kozak, 1984). In general, mature mRNA transcript is scanned from the capped 5' terminus until reaching the first AUG with an appropriate sequence context ($CC^{A}/_{G}C$<u>AUG</u>G), in which the underlined A is considered as '+1'. Comparison of sequences associated with known start codons suggests that -2 is a purine (A/G) and +4 is a G nucleotide for the efficient initiation of the translation process (Sakai et al., 2001). It means that if the sequence proximal to the start codon is not optimal, the ribosome could highly likely bypass the locus and continue to scan the mRNA for alternative start sites. Ribosome scanning can have an impact on translational activities, resulting in alterations of protein expression (Faure et al., 2016). Since we are unclear which of the prospective start codons is most likely to act as the major translation start site, to be on a safe side, we targeted the region downstrean the third start codon.

After that, we designed a single-stranded oligodeoxynucleotide (ssODN) template according to the target site of sgRNA we discussed above. We designed a ssDNA donor with asymmetric homology arms. Moreover, a silent/blocking mutation of PAM was introduced to inhibit the potential Cas9 re-cutting event. However, HDR is still in a low frequency (**Section 4.4.3**). We detected 2 HDR events and 11 NHEJ events out of 96 well-grown cell clones from 9 x 96 well plates in total. Most of these mutants were repaired by NHEJ pathway, resulting in the formation of small indels (1-3 bp) at the DSB locus. Note that this NHEJ-strategy is plausible only when the in/dels

are not triplets, otherwise the function of the protein could be still retained (Lei S et al., 2016).

Therefore, to precisely insert a premature stop codon into the genome of ST6GAL1, HDR is desired. To boost HDR frequency, several methods have been developed. For example, we could inhibit NHEJ rate or upregulate HDR pathways. For example, we could suppress KU70, KU80, or DNA ligase IV (Chu et al., 2015). We could control timing of CRISPR/Cas9 components delivery using cell synchronization (Lin et al., 2014). We could use cleavable donor DNA plasmid by flanking the donor vector with dual sgRNA recognition site (Zhang et al., 2017a). Tethering the DNA donor template to Cas9-guide RNA ribonucleoprotein (RNP) at the DSB site has been employed successfully to improve the HDR efficiency, without compromising the cell growth and cell different behaviours (Aird et al., 2018) Alternatively, 'cold shocking' cells at 32°C (1-2 days) following transfection could improve HDR frequency, which has been validated in induced pluripotent stem cells (Guo et al., 2018).

Apart from NHEJ and HDR events occurring in ΔCD clones, we found potential MMEJ events in ΔCD clones. More specifically, F6 and G2 exhibited a sgRNA-induced 55 bp and 8 bp deletion, respectively. The deletion flanking the break site was probably due to microhomology-mediated end joining (MMEJ) which was described in **Section 4.1.1.1.** MMEJ relies on 5-25 bp exposed microhomologous sequence to mediate the end-joining events at the DNA break site. Moreover, research suggests that MMEJ is not a back-up mechanism: it can operate when NHEJ and HR repair mechanism are intact (McVey and Lee, 2008).

After the validation of the mutation events in all the ΔCD cells, we sent all successfully identified cells for sequencing. The sequencing results showed strong signals and no evidence of heterozygous deletion. Therefore, we confirm them as homozygous mutates. In future, we could further confirm their homozygosity by performing two sequential rounds of PCRs. More specifically, the products from the spanning PCR (**Figure 4. 29**) can be subject to the second PCR with one internal primer and one external primer used in the spanning PCR. The internal primer is the one targeting the expected mutation site. -/-CD will not produce any amplicons in the second round of PCR, while WT and -/+CD will produce an amplified product.

In summary, these two edited NPTr cells lines harbouring ST6GAL1 precise editing are successfully established. In the next chapter, the functional study regarding cell susceptibility to IAV will help us better understand the molecular mechanisms involved in ST6GAL1 deficiency.

4.6 Conclusion

In this chapter, we have created two pig trachea cell models harbouring deficient ST6GAL1 expression using the CRISPR/Cas9 system. In the first strategy, two untranslated exons and the associated predicted promoter of airway-enriched transcripts of ST6GAL1 were targeted. The obtained cell line was named as ΔP. In the second approach, functional knockout of ST6GAL1 was generated in NPTr cells, and this model was named ΔCD. In the subsequent chapter, five -/-ΔP and 13 -/-ΔCD cells are subject to phenotypic and viral assessment.

5 Analysis of sialic acid expression and IAV infectivity

5.1 Introduction

Sialic acid (SA) is a nine-carbon monosaccharide typically found at the outermost positon of cell-surface glycan chains. The conjugation between SA and terminal carbohydrate chains can be in α 2,6- or α 2,3-glyosidic linkage, and is important in viral tropism and binding specificity. Human IAV preferentially binds to N-acetyl-neuraminic acid α 2,6-Galactose (SA α 2,6-Gal) receptor, and avian IAV predominantly binds to SA α 2,3-Gal receptor. The planar structures of these two receptors are demonstrated below (**Figure 5. 1**).

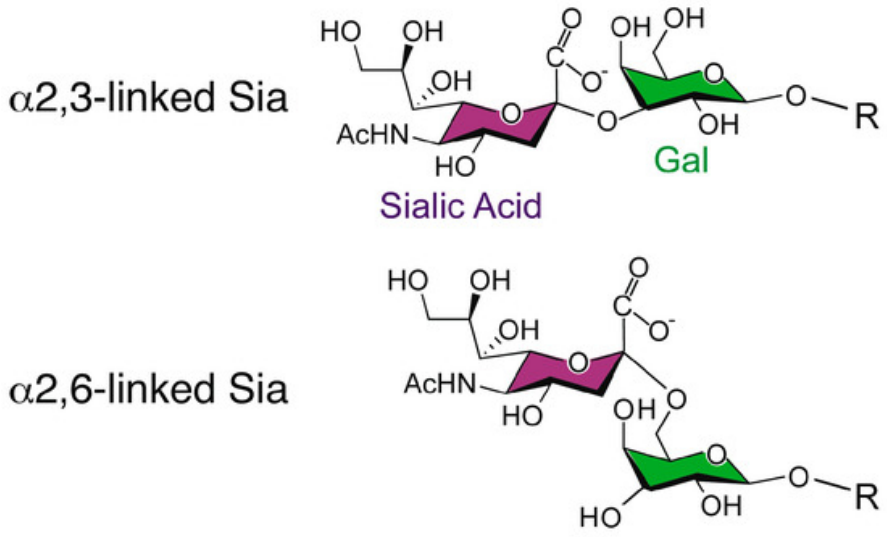


Figure 5. 1 The structure of SA α 2,6-Gal and SA α 2,3-Gal. The image was original from Lauren Bryd-leotis (Byrd-Leotis et al., 2017). The terminal position of sialic acid (SA or Sia) can bind to the penultimate position of Galactose (Gal) on the carborhydrate susbtrates with two types of linkages, α 2,3-linkage (upper) and α 2,6-linkage (bottom). The linkages are important in the specificty of viral binding. Please also note that the three-dimensional structure of α 2,6-linked SA is 'umbrella-like' and α 2,3-linked SA is 'cone-like' (Wilks et al. 2012).

Lectin staining can be performed to detect the terminal SA residues. *Sambucus Nigra* (SNA), a lectin isolated from elderberry bark, has been widely used to recognise α 2,6-SA (Fischer and Brossmer, 1995). The other one is *Maackia amurensis* lectin (MAL), which is used to detect α 2,3-SA. There are two forms of MAL, MAL I and MAL II. They differ by the conjugation between β 1,4- or 1,3- linked Gal and the next 'C' site. MAA I is able to identify non-SA residues and may cross-react with α 2,6-SA. MAA II can bind to a particular carbohydrate substrate containing an α 2,3-SA, and is thus selected in this experiment (Kuchipudi et al., 2009). *Concanavalin* A (ConA) lectin is used as a positive control: it can recognise α mannose known as 'core oligosaccharide' in many serum and glycoproteins on cell surfaces, which we believe is not depended on ST6Gal1 expression.

The newborn pig trachea (NPTr) cell line is a suitable cell model in IAV studies thanks to its susceptibility spectrum and SA-containing receptor distributions. It is a type of non-carcinoma and non-transformed cell line established from a 2-day-old piglet. The NPTr cell line appears to grow uniformly as a population of epithelial-like cells with oval nuclei. During subsequent passages, cells usually remain tightly packed in the form of monolayers, and seldom vary morphologically (Ferrari et al., 2003).

With this cell model, we have used the CRISPR/Cas9 system to precisely engineer ST6GAL1 (**Chapter 4**). Two types of homozygous edited NPTr cells were generated. One attempt was to alter the transcriptional profile of ST6GAL1 in the respiratory tracts, by deleting a 5' TSS region in the dominant transcripts expressed in the airway. We anticipated that in this case the synthesis of α 2,6-sialylated glycan structures was altered exclusively on the respiratory tracts. The resulting -/- edited cells C1, D12, F4, F7, and a3 were named as -/-ΔP1, -/-ΔP2, -/-ΔP3, -/-ΔP4, and -/-ΔP5,

respectively. The other attempt was a functional deletion of ST6GAL1 in NPTr cells, resulting in less engagement of the carbohydrate substrate in catalysing sugar binding. The resulting -/- edited cells were defined as -/-ΔCD. -/-ΔCD1, -/-ΔCD4, -/-ΔCD5, -/-ΔCD7, and -/-ΔCD11 were selected for the phenotype assessment.

In this section, we evaluated the efficiency of these strategies by the analysis of α 2,6-SA expression levels in edited NPTr cells, as well as assessing the cell susceptibility to three IAV strains.

To assess IAV replication, viral genome copy numbers and viral growth titres are used to determine viral replication. We use nucleoprotein (NP) as a parameter to evaluate IAV replication. NP, as a structural protein, can interact with viral genome and viral RNA-dependent RNA polymerase (RdRp) to form ribonucleoprotein (RNP). NP is an important factor for viral RNP nuclear transport and assembly, which is essential for the efficient viral infection (Portela and Digard, 2002).

5.2 Results

5.2.1 Viability of WT and edited NPTr cells in viral growth medium

The viability of cells cultured in viral growth medium (VGM) was determined using the CellTitre-Glo® Luminescent Cell Viability Assay (Cat# G7570, Promega) as described previously (**Section 2.1.1.2**). Given that the quantity of ATP luminescence was proportional to the number of viable cells in culture, we measured ATP luminescent signals to evaluate the viability of wild-type (WT), -/-ΔP2 and -/-ΔCD5 NPTr cells cultured at 0 h, 8 h, 24 h and 48 h. The luminescent signal of all samples was then normalised to the medium-only control. Subsequently, these measurements were plotted relative to the one produced by WT cells (time=0 h) (**Figure 5. 2**). The result implies that there are no significant differences of ATP productions among WT, -/-ΔP and -/-ΔCD cells, which in turn suggests that VGM culture does not have significant effects on the viability of WT and edited NPTr cells.

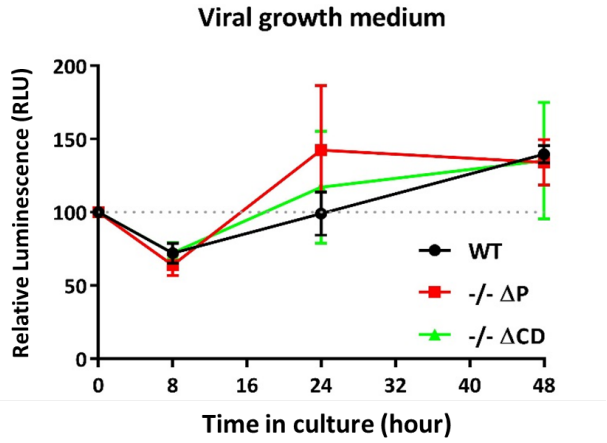


Figure 5. 2 The viability of NPTr cells in viral growth medium. The relative luminescence (RLU) produced in NPTr cells on 0 h, 8 h, 24 h and 48 h culture was plotted (y-axis). Two technical replicates were performed in two biological repeats (n=2).

5.2.2 Cell surface sialic acid expression levels

To detect α 2,6-SA and α 2,3-SA expression levels on NPTr cells, fluorescein-labelled lectin-based assay with SNA and MAL II were performed, respectively (**Section 2.3.3.2**). This was followed by the visualisation using confocal microscopy and the quantification using flow cytometry (**Section 5.2.2.3**).

Throughout this section, ConA was used as a positive control (+ve) since we expected that ConA lectin binding would be less influenced by ST6Gal1 activity. There was no negative control to take into account the non-specific binding of the lectin in cells. A proper negative control could be cells without SA on the cell surface, which could be achieved by treating cells with sialidase enzyme (Fischer and Brossmer, 1995). However, this had not been performed due to the time constraint.

5.2.2.1 Sialic acid expression of MDCK and MDCK-SIAT1 cells by lectin stain

As a preliminary, we observed α 2,6-SA and α 2,3-SA expression levels in MDCK cells and MDCK-SIAT1 cells through SNA and MAL II lectin fluorescent signals, respectively. It has been shown that MDCK-SIAT1 cells have an enhanced level of α 2,6-sialyltransferase than conventional MDCK cells, and thus increased α 2,6-SA expression (Lin et al., 2017). In our study (**Figure 5. 3)**, the positive control, ConA lectin staining by confocal imaging showed strong fluorescent signal of SNA lectin in MDCK and MDCK-SIAT1 cells as expected. Moreover, the confocal imaging of MAL II lectin binding showed moderate level of fluorescent density in MDCK and MDCK-SIAT1 cells. Importantly, it was noted that SNA lectin fluorescent signal in

MDCK-SIAT1 cells was obvious higher than that in MDCK cells, which is consistent with the previous report (Lin et al., 2017).

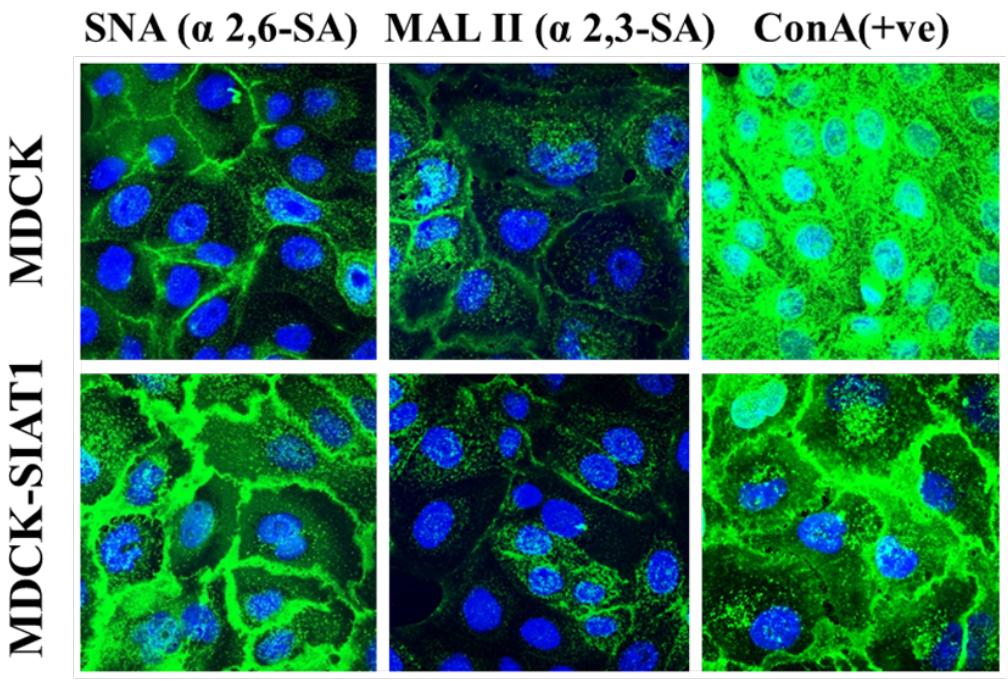


Figure 5. 3 Analysis of α 2,6-SA expression in MDCK and MDCK-SIAT1 cells. Magnification: 64 x. They are representative of two independent staining (n=2).

5.2.2.2 Evaluation of cell surface sialic acid expression levels

All the edited cells in this study were developed through clonal expansion, and it was unclear whether surface glycan expression was altered during cell clonal expansion. For this reason, we performed SNA, MAL II and ConA lectin-based staining in pooled WT cells and single cloned WT cells. As seen in **Figure 5. 4a**, for each staining experiment, pooled WT cells and single cloned WT cells displayed no obvious difference of fluorescent signals.

Furthmore, flow cytometry analysis was performed to measure α 2,6-SA expression in both pooled WT cells and single cloned WT cells (**Figure 5. 4-b**). A sample, consisted of unstained cells, was included to set positivity of fluorescent signals. The fluorescent intensity histogram showed that there were no obvious differences in mean fluorescent intensity (MFI) between the WT cell population and the cloned cells (**Figure 5. 4-b**). Therefore, we conclude that clonal expansion had little effects on the cell surface α 2,6-SA expression level in our experiment.

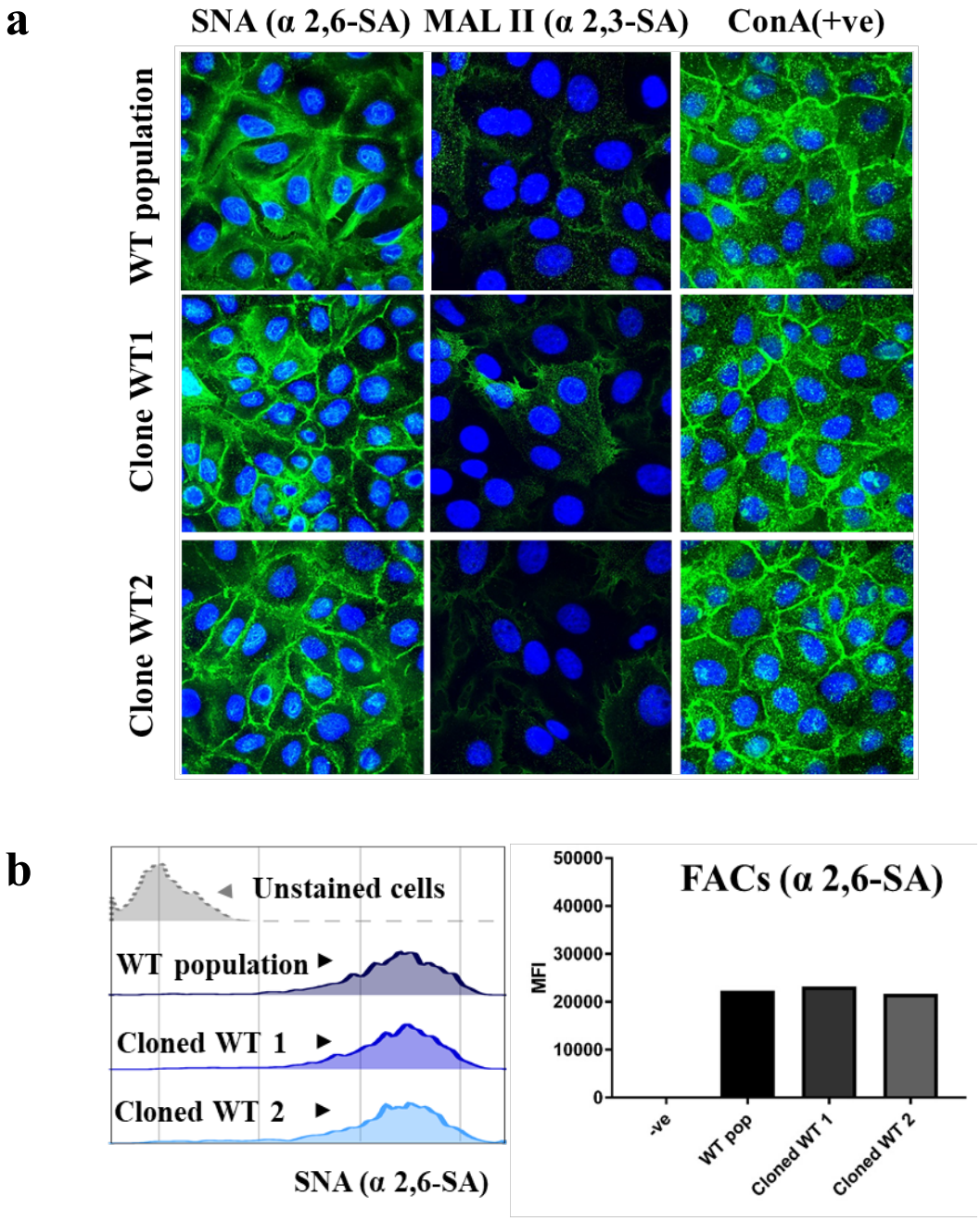


Figure 5. 4 Analysis of α 2,6-SA expression in cloned NPTr cells. (a) Fluorescein-labelled SNA (for α 2,6-SA, left), MAL II (for α 2,3-SA, middle) and ConA (+ve, right) lectin staining showed similar fluorescent signals in WT and cloned cells. (b) The left panel was the fluorescent intensity histogram (B530, y-axis) versus cell number (x-axis). The negative control (unstained cells) was used to set gating. The right panel is the flow cytometry quantification of α 2,6-SA expression levels among three samples. Magnification: 64 x. There were no biological replicates (n=1).

5.2.2.3 Evaluation of sialic acid expression in ST6GAL1 ΔP cells

To determine the expression and distribution of SA in edited NPTr cells, cell samples were stained with SNA, MAL-II, and ConA lectin. As can be seen in **Figure 5. 5,** ConA lectin staining (bottom panel) was visualised by confocal microscopy and showed a similar level of fluorescent signals in -/+ΔP (heterozygous edited), -/-ΔP cells (**Chapter 4**), and WT cells. Regarding fluorescein-labelled SNA lectin staining (upper panel), the fluorescent signals in -/+ΔP and -/-ΔP cells appeared reduced than those in WT cells, indicating that the genetic manipulation of ST6GAL1 in NPTr cells caused a reduction of surface α 2,6-SA expression.

Interestingly, the fluorescein-labelled MAL II lectin staining (middle panel) showed that the α 2,3-SA expression level in -/+ΔP and -/-ΔP cells was higher than that in WT cells. This may suggest that α 2,6-SA and α 2,3-SA are in equilibrium to each other on cell surfaces: the reduction of one may result in the increase of the other.

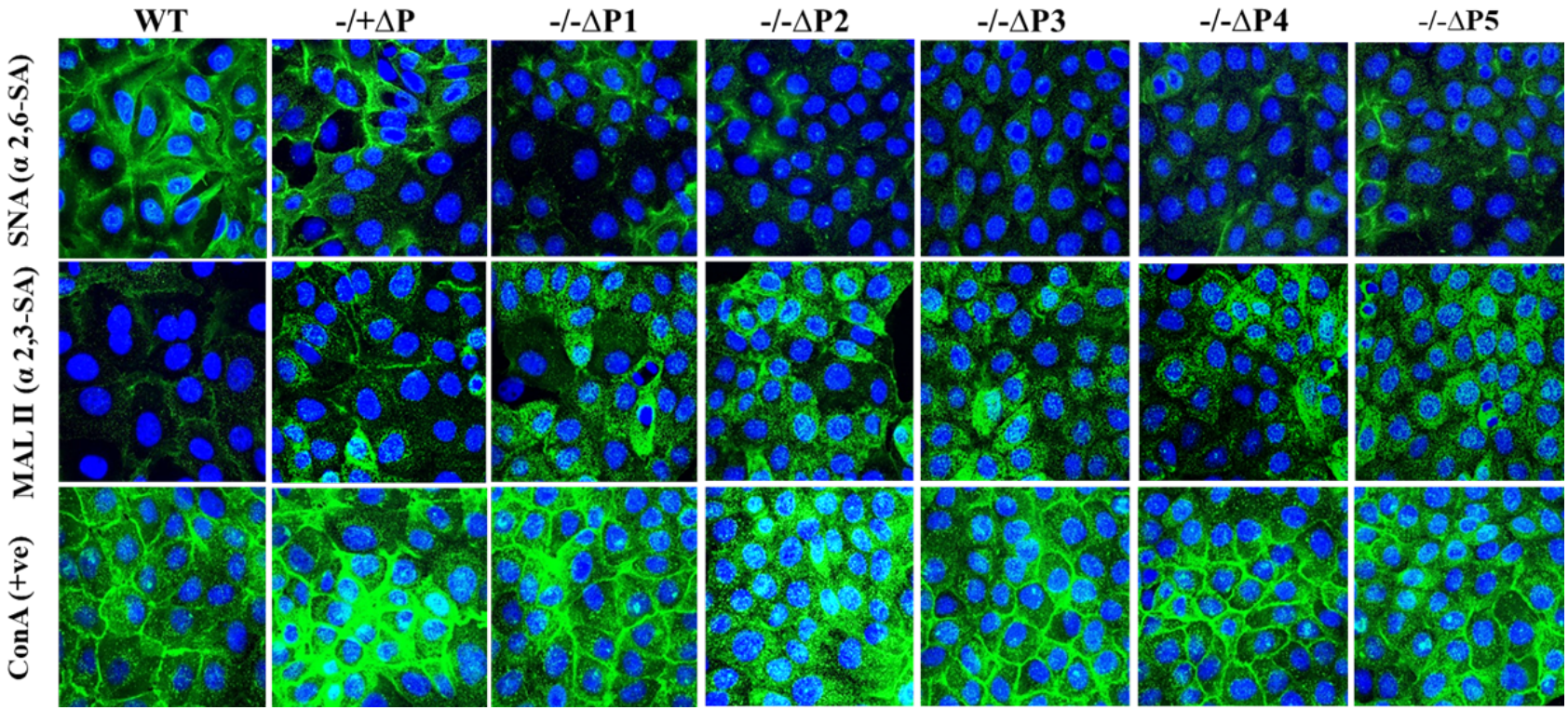


Figure 5. 5 Expression and distribution of sialic acid in ΔP cells. Cells were seeded on coverslips and stained with fluorescein-labelled SNA (upper), MAL II (middle), and ConA (Bottom) lecints, followed by confocal imaging. Images were taken by confocal microscopy. The fluorescent green is SA distribution and blue is cell nuclei (DAPI). Magnification: 64 x.

5.2.2.4 Cell surface sialic acid expression in ST6GAL1 ΔCD cells

We detected the α 2,6-SA expression level in 13 clones of -/-ΔCD cells using SNA staining. Data shown here is representative of five out of 13 (**Appendix 13**). As shown in **Figure 5.** , the confocal imaging of SNA lectin staining indicated that α 2,6-SA expression level appeared lower than that in WT cells, which we believed was due to the functional domain deletion of ST6Gal1 protein in -/-ΔCD cells (**Chapter 5**). The MAL II lectin staining by confocal microscopy displayed moderately stronger fluorescent signals in -/-ΔCD cells than in WT cells, providing more evidence that the expression of α 2,6-SA is closely related to α 2,3-SA.

It is worth noting that the positive control, ConA lectin binding, turned out to be stronger in most of the -/-ΔCD cells than in WT cells, which was unexpected. We shall address this in the discussion section (**Section 5.3**).

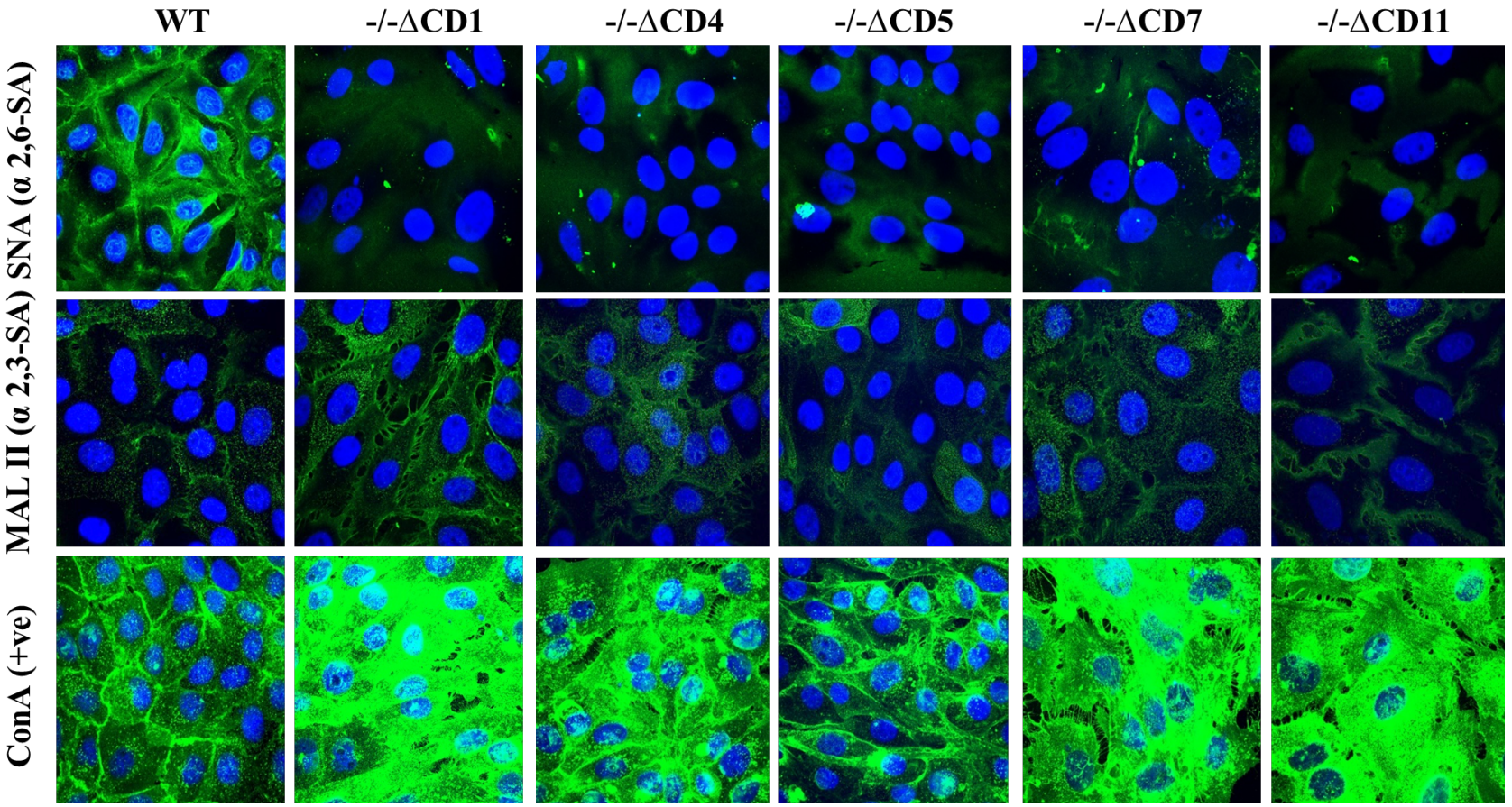


Figure 5. 6 Expression and distribution of sialic acid in ΔCD cells. Cells were seeded on coverslips and stained with fluorescein-labelled SNA (upper), MAL II (middle), and ConA (Bottom) lecints, followed by confocal imaging. Images were taken by confocal microscopy. The fluorescent green is SA distribution and blue is cell nuclei (DAPI). Magnification: 64 x

5.2.2.5 Quantification of α 2,6-SA by flow cytometry in -/-ΔP and -/-ΔCD cells

SNA-labelled fluorescent intensities of WT cells, -/+ΔP, -/-ΔP 1, 2, 3, 4, 5, and -/-ΔCD 4, 5, 11 cells were quantitatively analysed by flow cytometry as described previously (**Section 2.3.3.2**). MFI was measured to determine the central tendency of SNA fluorescent signal, which correlated the α 2,6-SA expression level on the cell surface. As shown in **Figure 5. 7**, cells unstained with SNA lectin (negative control, -ve) were included to set a gating in the fluorescent histogram graph. In the flow cytometric histogram (**Figure 5. 7-a**), compared with those in WT cells, we observed that the peak of MFI was left shifted in SNA-labelled -/-ΔP cell 1-5, and the peak shift further towards the left in -/-ΔCD cell 4, 5 and 11. There was no significant difference of MFI in -/+ΔP cells versus in WT cells.

Accordingly, the histogram bar graph (**Figure 5. 7-b**) showed the MFI value produced by SNA binding in each sample. Each MFI was then normalised to the levels of the negative control. Compared to that in WT cells, the MFI value in -/-ΔP was significantly reduced; and the MFI value of -/-ΔCD cells had a further significant reduction. The quantification analysis was consistent with the visualisation by lectin staining (**Section 5.2.2.3** and **Section 5.2.2.4)**, indicating that α 2,6-SA expression levels were reduced in both -/-ΔP and -/-ΔCD NPTr cells.

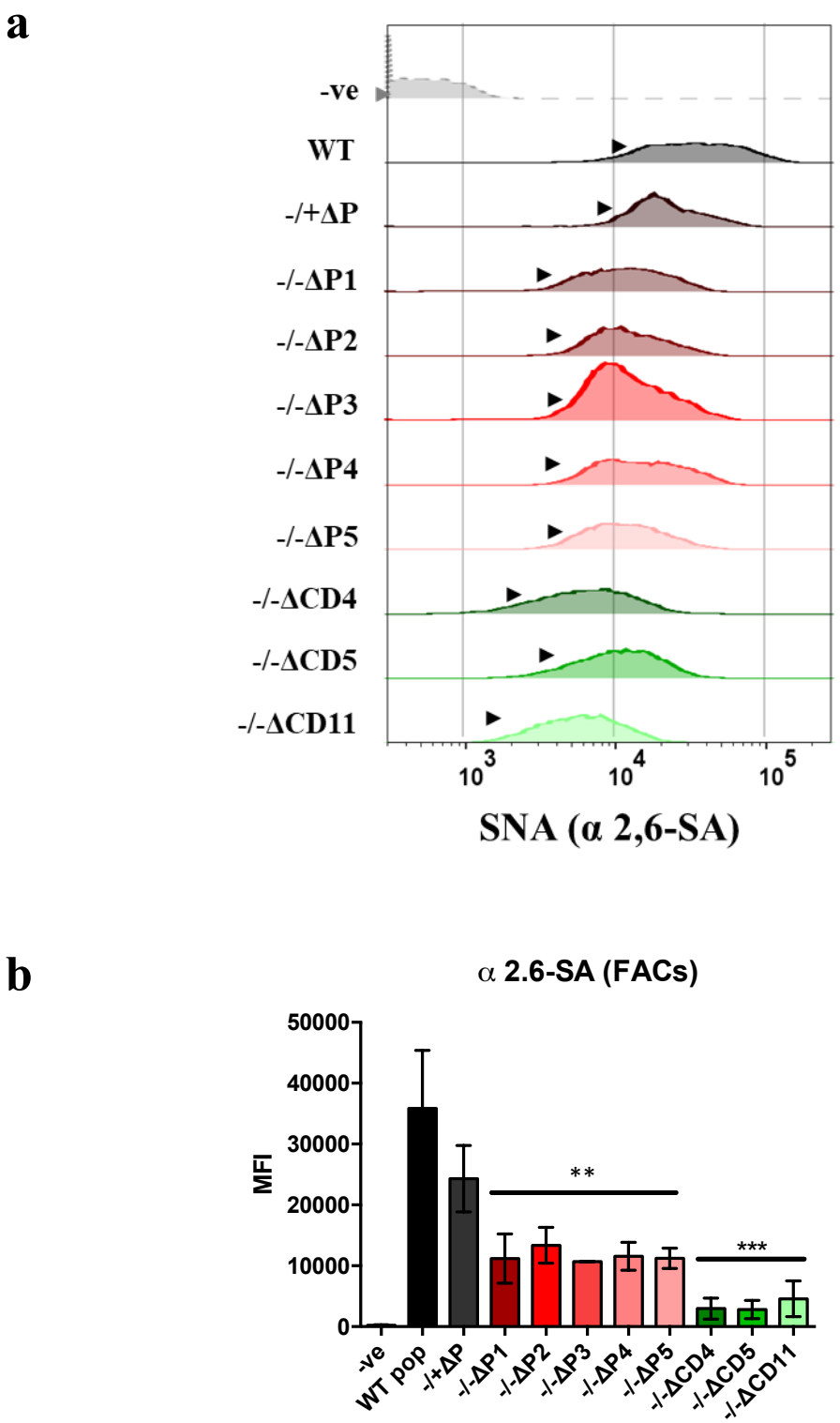


Figure 5. 7 Quantification of SNA binding by flow cytometry. (a) The histogram for SNA measured by flow cytometry. (b) The histogram bar chart for quantitative MFI of α 2,6-SA in -/-ΔP and -/-ΔCD cells. Statistical analysis was performed in three independent experiments (n=3). Values were mean ± SEM Significantly different values (*p<0.05, ***p<0.001) were calculated by one-ANOVA analysis.

In parallel, we determined the surface α 2,3-SA expression level on MAL II-stained cells using flow cytometry as described previously (**Section 2.3.3.2**). As shown in **Figure 5. 8** (upper panel). Compared to the MFI of MAL II in WT cells, the peak was right shifted in edited cells, indicating the increased level of α 2,3-SA binding in the edited cells. As can be seen in the bottom panel (**Figure 5. 8**), the statistical analysis revealed the MFI ratios of MAL II lectin were increased in edited cells accordingly. The results suggested an increased expression levels of α 2,3-SA in ST6GAL1 edited cells. There were two technical replicates, and three biological replicates in each group due to the time constraint.

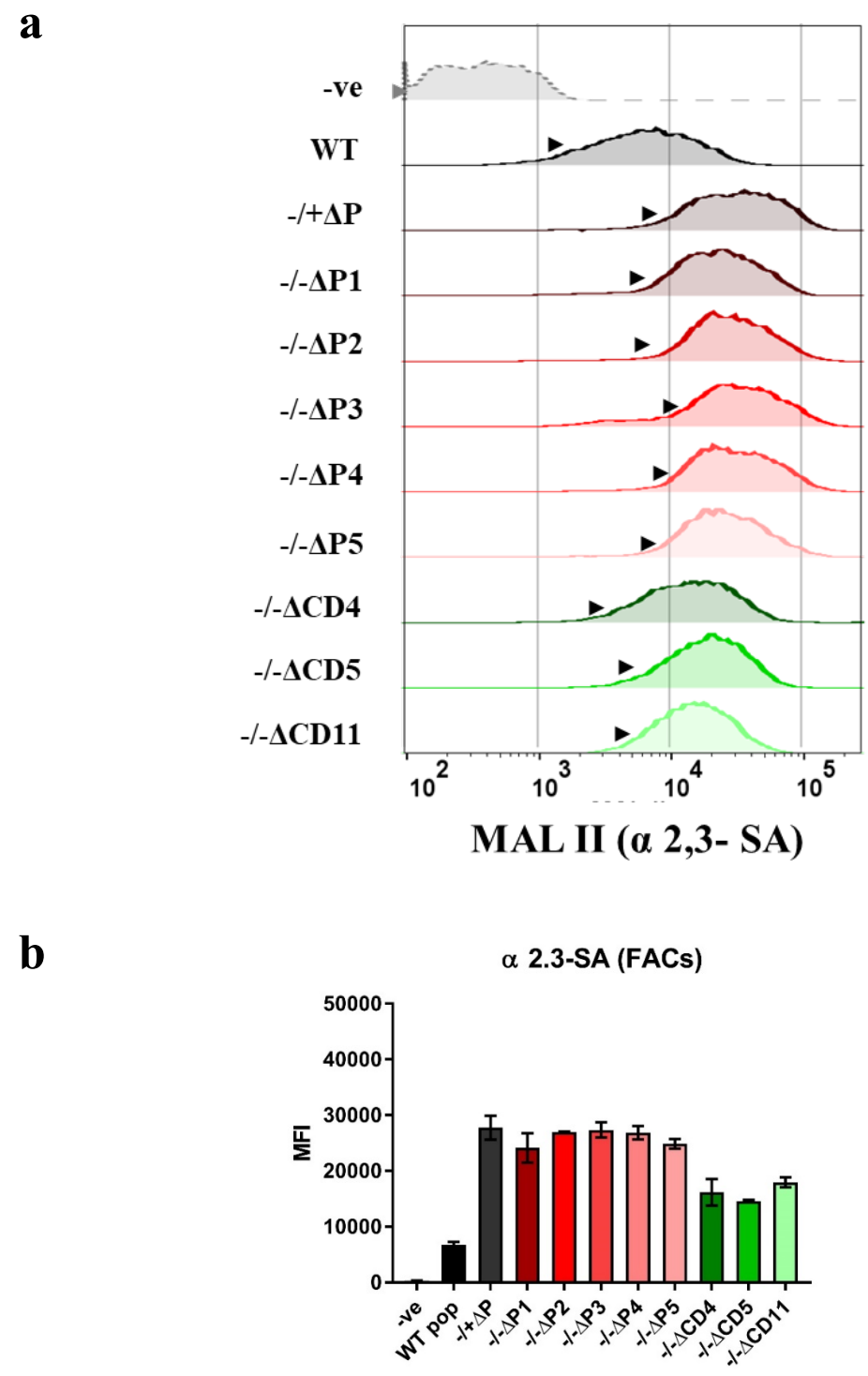


Figure 5. 8 Quantification analysis of MAL II binding by flow cytometry. (a) The histogram for MAL II measured by flow cytometry. (b) The histogram bar chart for quantitative MFI of α 2,3-SA in -/-ΔP and -/-ΔCD cells. Statistical analysis was performed twice (n=2).

5.2.3 Single-cycle infection of IAV

5.2.3.1 NP expression analysis by Western blot

5.2.3.3.1 PR8 and Swine87 viral infection

Now that the SA distribution on WT, -/-ΔP, and -/-ΔCD NPTr cells has been evaluated, the next question is whether editing ST6GAL1 has an effect on viral replication. To this end, the levels of intracellular nucleoprotein (NP) production were quantified in a single-cycle infection using A/Puerto Rico/8/1934 (PR8) and A/Swine/England/163266/1987 (Swine87) viruses, respectively. PR8 virus is a lab-adapted IAV strain known to infect host cells by binding to both α 2,6-SA and α 2,3-SA-containing receptors (private communication with John McCauley lab). Swine87 was chosen to present triple-assorted viral strains. WT cells, and three -/-ΔP clones (2, 4, 5), and three -/-ΔCD clones (4, 7, and 11) were infected with PR8 and Swine87 virus, respectively, at an M.O.I of 0.2 for 8 hours post infection (h.p.i). NP production was then quantified by western blot analysis.

As shown in **Figure 5. 9**, mock-infected cells had undetectable NP productions. The difference between WT cells and -/-ΔP cells in terms of NP productions during PR8 (**a**) and Swine87 (**b**) viral infection was insignificant. Furthermore, compared to WT cells, -/-ΔCD cells exhibited similar levels of NP productions during PR8 viral infection. However, NP expression following Swine87 infection was found to be significantly lower (P-value<0.01) in -/-ΔCD cells.

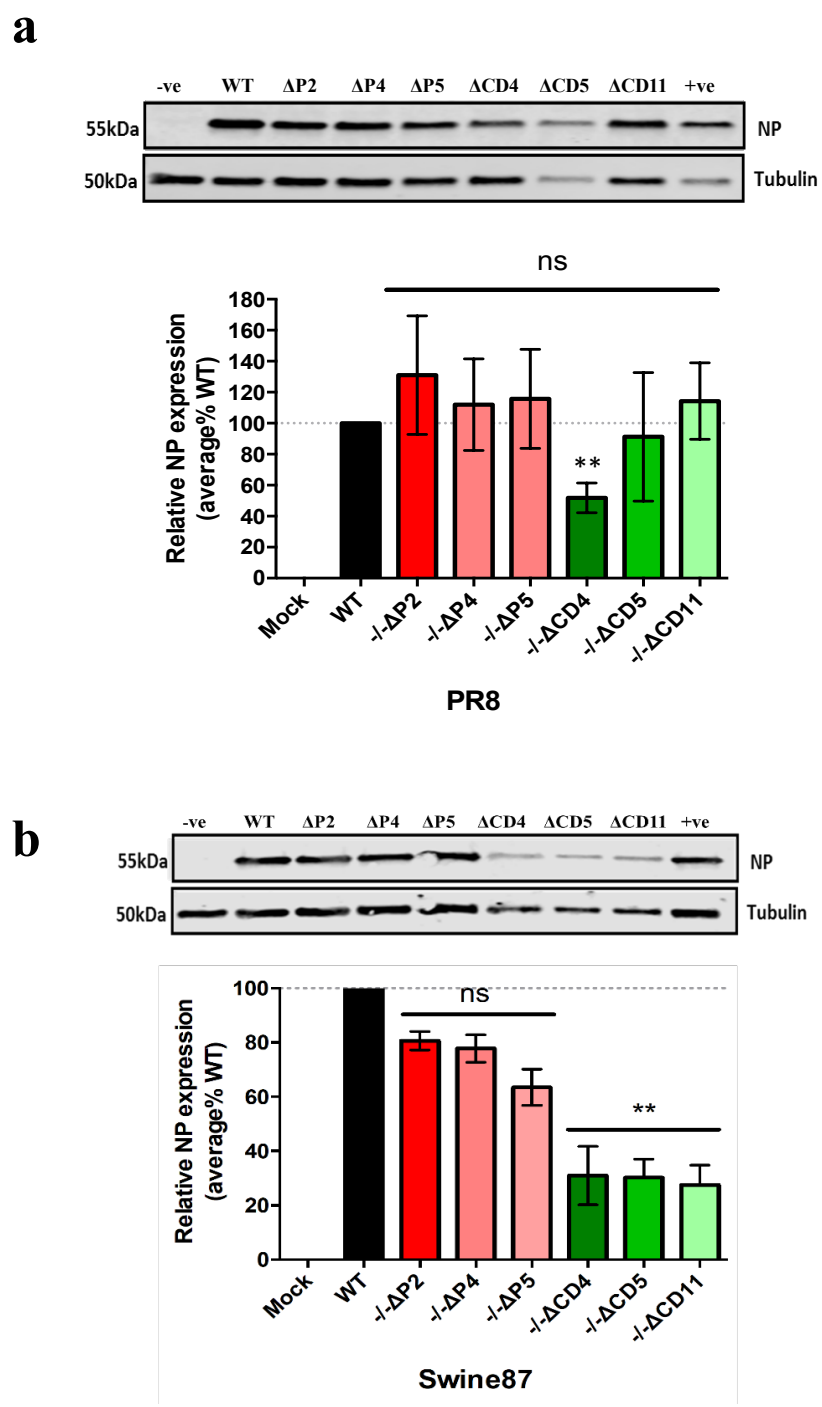


Figure 5. 9 NP replication at 8 h.p.i with PR8 and Swine87 by western blot. NPTr cells at 8 h.p.i by PR8 (a) and Swine 87 (b) with an M.O.I of 0.2 were harvested. NP was determined by western blot using anti-NP antibody. The histograms below each blot were the quantitate analysis by the normalisation to Tublin. All error bars were calculated with S.E.M and the P-value was determined by one-way ANOVA analysis. **P-value=0.0073< 0.01. ns: not significant.

5.2.3.3.2 Cal04 viral infection

A/California/04/2009 (H1N1) (Cal04) virus was chosen as a representative pdm09 virus. Cal04 virus has been shown to preferentially bind to α 2,6-SA receptor. It is important to evaluate Cal04 viral infectivity since we would like to reduce pdm09 replication in pig cells.

The ability of Cal04 virus to replicate in ST6GAL1 edited cells was evaluated by a single-cycle infection using an M.O.I of 0.5. Infected cells were harvested at 24 h.p.i, as NP production was undetectable at 8 h.p.i (see **Appendix 14**) and a longer incubation was required. As shown below (**Figure 5. 10**), mock-infected cells had undetectable NP production. The level of NP production in -/-ΔP cells was significantly lower than that in WT cells. Furthermore, NP production was even lower in -/-ΔCD cells. In fact, we did not observe any NP productions by western blot; and only detected a very low level of NP production by densitometry.

Taken together, our results indicate that, compared with WT cells, -/-ΔP cells were less susceptible to Cal04 viral infection and that -/-ΔCD cells are less susceptible to Swine87 and Cal04 viral infection.

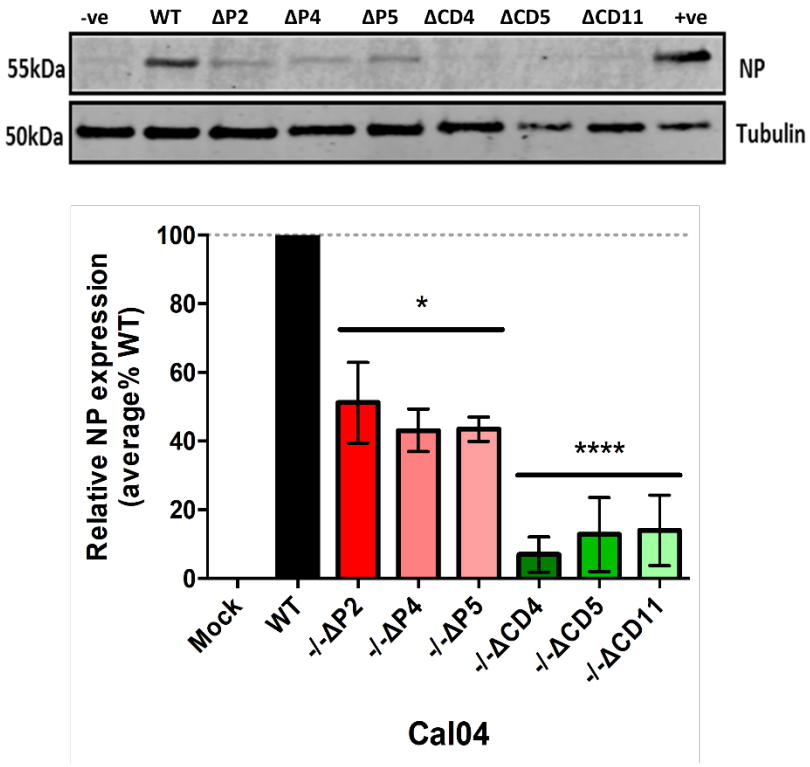


Figure 5. 10 NP replication at 24 h.p.i with Cal04 by western blot. Western blot analysis showed the viral NP expression in infected cells. All error bars were calculated with S.E.M and P-value was determined by one-way ANOVA analysis 24 h.p.i by Cal04 virus at M.O.I of 0.5. The histogram bar was the quantitate analysis by normalisation to Tublin. The P-value < 0.05(*), and 0.0001 (****).

5.2.3.2 NP expression visualisation in infected NPTr cells by immunofluorescence

Next, we used the immunofluorescence assay (IFA) to visualise intracellular NP expression. To this end, WT, -/-ΔP and -/-ΔCD cells were seeded on coverslips and infected with PR8, Swine87, and Cal04 viruses, respectively (**Section 2.2.2.2**). The infected cells were fixed and permealised at 8 h.p.i (PR8 and Swine87 viruses) or 24 h.p.i (Cal04 virus), and stained with the anti-NP antibody (**Section 2.2.2.4**). The stained cells were then visualised by fluorescent microscopy, which allowed us to quantitatively measure the numbers of fluorescent-labelled infected cells in different groups. These numbers were plotted relative to that of WT cells. The uninfected cells (mock-infected) were used as a negative control (-ve).

As shown in **Figure 5. 11**, mock-infected cells showed no fluorescent signals in WT, -/-ΔP and -/-ΔCD cells. Quantitative analysis of the fluorescent signals in PR8-infected WT, -/-ΔP and -/-ΔCD cells showed no significant differences among them at 8 h.p.i Regarding the infection of Swine87 virus, WT and -/-ΔP cells presented similar infectivity. However, if we measured susceptibility by the percentage of NP positive cells, -/-ΔCD cells were approximately 50 percent less susceptible to Swine87 than WT cells. For Cal04 viral infection, -/-ΔP and -/-ΔCD cells were 20% and 70% susceptible to the virus than WT cells, respectively.

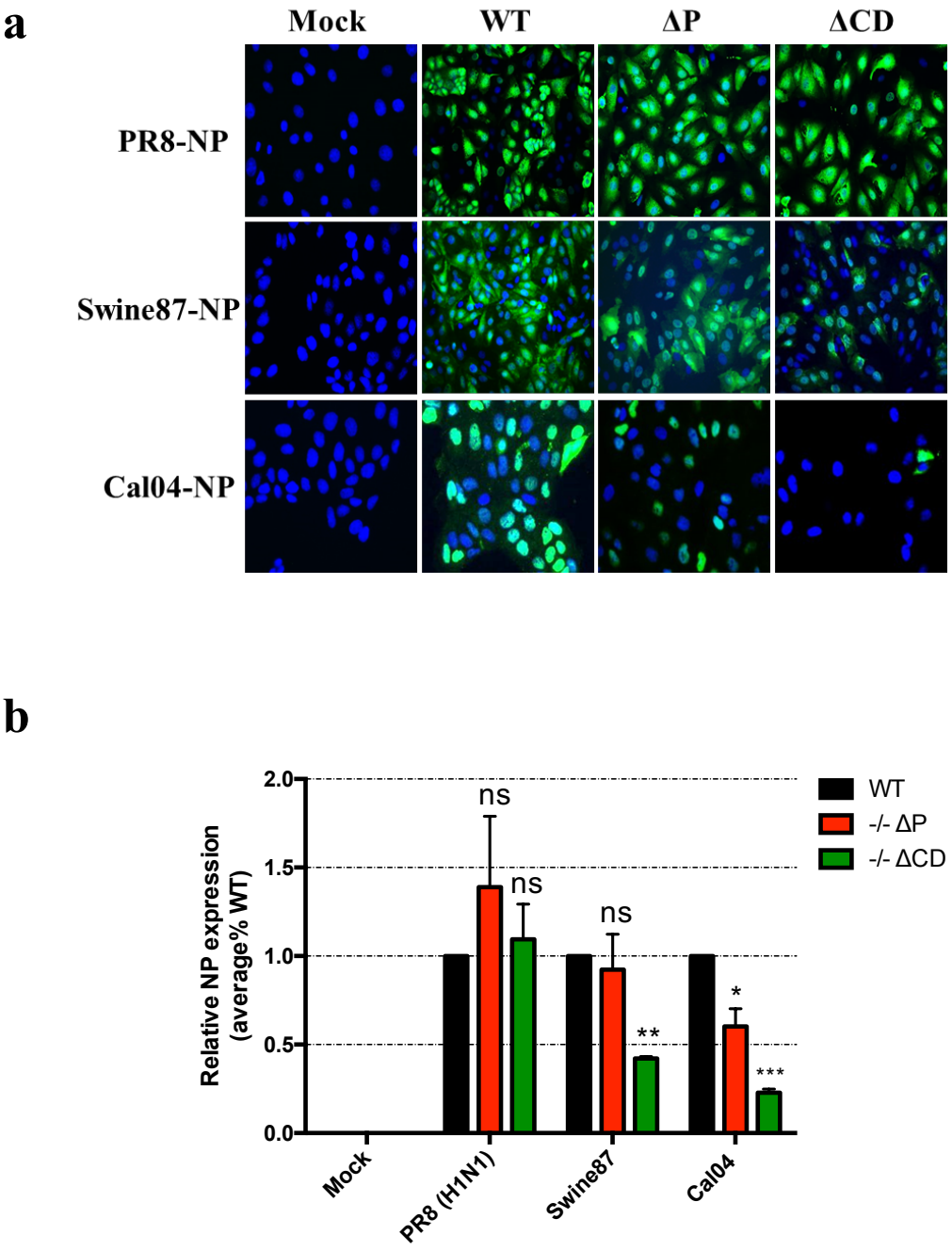


Figure 5. 11 NP expression by immunofluorescence assay. (a) Cell on the coverslips were infected and harvested at 8 h.p.i by PR8 and Swine87 viral infection, and 24 h.p.i by Cal04 viral infection prior to IFA using NP monoclonal antibody. Nuclei was counterstained with DAPI dye. (b) The values of viral NP in -/-ΔP and -/-ΔCD cells was normalised to that in WT cells. The quantification was based on cell counting in multiple micrographs from three independent experiments (n=3). * P<0.01 and **P <0.001. mean±SEM. Microscopy:100x.

5.2.3.3 NP expression quantification in IAV infected NPTr cells by flow cytometry

In this experiment, we infected cells using the same conditions as the **Section 5.2.3.2,** but quantified the NP expression by flow cytometry. As shown in **Figure 5. 12a**, cells in the negative control were mock infected and then stained. This negative control was used to record the non-specific fluorescent signals (more than 99.9 percent of these signal was included). Then, NP fluorescent intensities of IAV-infected cells were compared to that of mock-infected cells. This was achieved by normalising the percentages of cells with fluorescent signals against that of the negative control. The resulting numbers were plotted in **Figure 5. 12a.** Furthermore, this experiment was repeated independently for three times, and average percentages of NP+ cells in different IAV-infected cell population were reported in the bar chart (**b**).

Regarding PR8 viral infection, the normalised percentage of cells with fluorescent signals was 43.7% in PR8-infected WT cells, indicating that 43.7% of the WT cells were viable and NP+. The numbers for -/-ΔP and -/-ΔCD were 43.0% and 41.4%, respectively. The results suggest that PR8 virus has similar infectivity to WT, -/-ΔP and -/-ΔCD cells. For Swine87 virus, the percentage of NP+ Swine87-infected WT cells was 32.5%, which was similar to the percentage for -/-ΔP cells, 33.6%. However, the percentage for -/-ΔCD was ~50% lower than that for WT cells. As for Cal04 viral infection, the percentage of NP+ cells in Cal04-infected -/-ΔP was ~20% lower than that for WT cells (37.3%) (**b**). For Cal04-infected -/-ΔCD cells, the number was ~60% lower than that for WT cells. Taken together, these data suggest that, compared to WT cells, -/-ΔP cells are less susceptible to Cal04 virus and that -/-ΔCD cells are less susceptible to both Swine87 and Cal04 viruses.

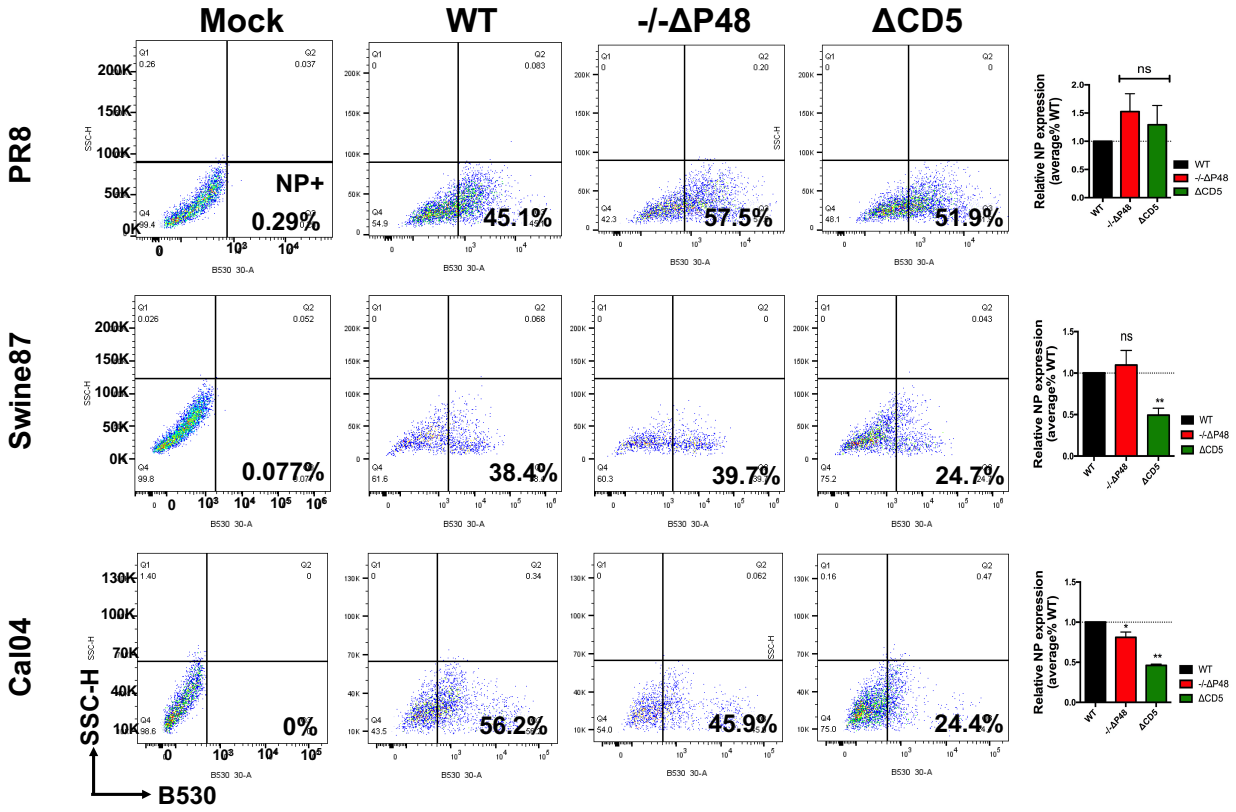


Figure 5. 12 NP expression quantification by flow cytometry. (a) Flow cytometry gating for the identification of NP positive cells. IAV-infected NPTr cells (same conditions with IF) were harvested, fixed, permeablised, and stained prior to flow cytometry analysis. Representative results showed the percentage of NP positive cells in each gate, which was proportional to the ratio of IAV-infected cells. The NP fluorescent intensity in mock-infected cells was used to set a gating (black box) and was recorded no more than 0.1% NP expression. (b) The quantification analysis according to (a). Data represented the mean ± SEM of three technical replicates (n=3)

5.2.4 Multiple-cycle infection of IAV

5.2.4.1 Replication kinetics of PR8, Swine87 and Cal04 in WT, -/-ΔP and -/-ΔCD cells

To determine the ability of viral replication in WT, -/-ΔP and -/-ΔCD cells, viral growth kinetics of PR8, and Swine87 virus was assessed in MDCK cells, and Cal04 virus was assessed in MDCK-SIAT1 cells. In both cases, we performed multiple-cycle infection at an M.O.I of 0.001, and cell supernatants were harvested at 8, 24, and 48 h.p.i (**Section 2.2.2.3**). Mock-infected cells contained undetectable virus. The viral growth curve was shown in **Figure 5. 13a**, during multiple-cycle infection with PR8 virus, the viral growth in WT, -/-ΔP and -/-ΔCD cells were very similar, with viral titres reaching at 10^8 PFU/ml at 48 h.p.i. For multiple-cylce infection with Swine87 virus, viral growth reached 10^5 PFU/ml in WT cells at 48 h.p.i. Compared with WT cells at both 24 and 48 h.p.i, -/-ΔP cells presented 1-log higher viral titre, which was unexpected; -/-ΔCD cells presented 1.5-log lower viral titre. Regarding Cal04 viral replication during the multiple-cycle infection, the viral titre in WT cells reached 10^6 PFU/ml at 24 h.p.i, and 10^7 PFU/ml at 48 h.p.i. Compared to WT cells, -/-ΔP cells replicated with delayed growth kinetics, showing a 1.5-log lower viral titre at 24 h.p.i, and 1-log lower viral titre at 48 h.p.i; Cal04 virus replicated with markedly delay over the 48 h.p.i. The difference of viral titres reached 4-log lower viral titre at 24 and 48 h.p.i, compared with WT cells. To summarise, these results show -/-ΔP cells allow the efficient viral replication of Swine87, but less viral growth of Cal04; -/-ΔCD cells are capable of allowing both Swine87 and Cal04 viral replication.

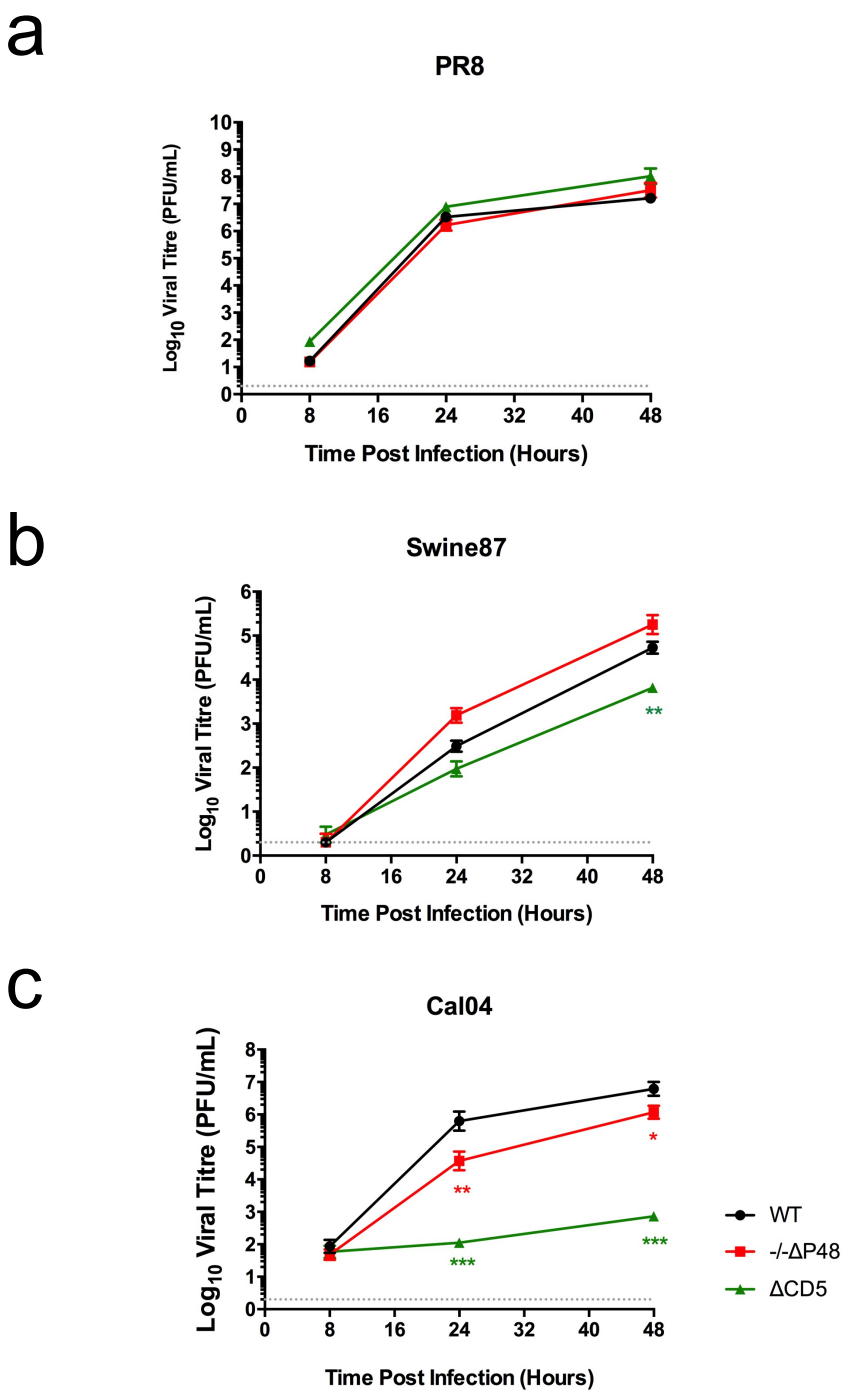


Figure 5. 13 Viral kinetics of PR8, Swine 87 and Cal04. Cell cultures were infected with viruses at an M.O.I of 0.001. Supernatants were harvested at 8, 24, and 48 h.p.i, and the viral titres were calculated as log10 PFU/ml in MDCK cells (PR8 and Swine87 viruses) or MDCK-SIAT1 cells (Cal04 virus). Plaque assay was performed and growth curves were plotted by Rute Pinto, a PhD student in Prof. Paul Digard group, the Roslin Institute. Values presented the standard mean±S.E.M (n=3). No statistically difference found.

5.3 Discussion

Firstly, we would like to discuss the distribution of the SA α 2,6-Gal and SA α 2,3-Gal receptors on cell surfaces. In humans, the SA α 2,6-Gal receptor is mainly distributed in the upper airway epithelium, the paranasal sinuses, the pharynx, the trachea, and the bronchi (Ibricevic et al., 2006), whereas the SA α 2,3-Gal receptor is predominantly distributed in non-ciliated bronchiolar cells (Shinya et al., 2006). In pigs, SA α 2,6-Gal receptor is predominantly expressed in the upper airway (e.g., trachea) (Kumlin et al., 2008), whereas SA α 2,3-Gal receptor is abundant in the lower airway (Zeng et al., 2013). In our experiment in pig trachea cells (the NPTr cells), we observed that SA α 2,6-Gal and SA α 2,3-Gal are both expressed on cell surfaces, and that of, SA α 2,6-Gal was roughly three times more abundant than SA α 2,3-Gal. In our study, we used lectin-staining assay to measure SA α 2,6-Gal and SA α 2,3-Gal expression levels, but it should be noted that we could achieve the same goal using mass spectrometry to measure the SA residues by digestion and fragmentation with specific sialidases (Zaia, 2010). As a positive control, we used ConA lectin to measure α mannose expression on cell surfaces. Compared with WT and -/-ΔP cells, a higher level of mannose expression was observed on the surface of -/-ΔCD cells. The homeostatic properties of the mannose receptors are known to be associated with cells' immune responses to foreign pathogens (Turner, 2003). Therefore, the functional knockout that generates the -/-ΔCD cells might cause a biological deficiency, which should be considered whether this would be a suitable solution for modelling host resistance in this case. However, these were only observations, more accurate quantification, such as flow cytometry, and protein-based assay, such as western blot, and the ELISA assay could be performed.

In addition, we observed that the ST6GAL1 deficient cells displayed reduced α 2,6-SA expression, and an increased level of α 2,3-SA expression, relative to WT cells. We conjecture that α 2,6-SA and α 2,3-SA receptors complement each other: the lack of one might cause the increase of the other. The way they depend on each other requires further investigations and is beyond the scope of this thesis.

An unexpected difficulty that arose during the quantitative analysis of NP production by flow cytometry. Due to the time constraint and project funding, the conditions and the concentration of anti-NP antibody have not been optimised. Anti-body NP was empirically used in 1:200 dilution. Therefore, to achieve a more accurate and robust result, optimization of antibody and collection of more cells (10,000 events) should be the first step before the quantitative analysis

5.4 Conclusion

Taken together, the manipulation of the ST6GAL1 gene rendered pig trachea cells less susceptible to IAV infection. In particular, -/-ΔP cells, which harbours ST6GAL1 perturbation in its transcriptional profile, exhibited less α 2,6-SA expression, and less susceptibility to Cal04 viral infection. -/-ΔCD cells, obtained by functional knockout of ST6GAL1, showed remarkably reduced α 2,6-SA expression, allowing them to be less susceptible to both Swine87 and Cal04 viral infection.

6 General discussion

6.1 Overview and discussion

Influenza is one of the top three economic diseases. An average of 600 people die of influenza each year in the UK. Pigs are reservoirs for human, avian, and swine IV strains since they have two types of sialic acid (SA)-containing receptors for viral recognition, SA α 2,6-Gal and SA α 2,3-Gal. This allows pigs to serve as intermediate hosts in the production of reassortant IAV strains, which facilitates cross-species transmission, and may even lead to epidemics and pandemics. In particular, the 'swine flu', pandemic (H1N1) 2009, isolated from pigs, has been widely circulating in human populations globally since 2009. Traditional vaccines and antiviral drugs are not fully sustainable due to the continuous emergence and re-emergence of IAV strains. Therefore, a growing body of literature has focused on host genetic factors involved in the IAV entry and infection, which may be useful for building host genetic resistance as a prophylaxis.

With the goal of building host resistance in pigs, we evaluate a strategy to reduce the expression of SA α 2,6-Gal receptor, which in turn could disrupt the virus-host interaction. To this end, we generated functional knockout of the ST6GAL1 gene, which regulates the configuration of SA α 2,6-Gal receptor. The resulting model, ST6GAL1ΔCD (ΔCD for short), exhibited dramatically reduced α 2,6-SA expression on cell surfaces. Furthermore, the edited cells were less susceptible to swine IAV strain, and resistant to a human IAV strain.

ST6Gal1 null mice showed severe deficiency of the humoral immunity (Nasirikenari et al., 2010), and suffered from the impaired B lymphocyte activation

(Hennet et al., 1998, Irons and Lau, 2018). Therefore, it is not ideal to generate animal models with the global loss of ST6GAL1. In our study, ΔCD cells exhibited normal viability (**Figure 5. 2**), but we were not clear whether the loss of ST6GAL1 would cause any immune insufficiency in these cells.

To be on the safe side, we put a more subtle approach to impair the configuration of SA α 2,6-Gal receptors. In particular, we targeted the 5' TSS region of two transcripts enriched in the respiratory tract, with the aim of disrupting the expression level of ST6GAL1 exclusively in the relevant tissues.

The difficulty of the aforementioned strategy was that we did not fully understand the transcriptional profiling of the pig ST6GAL1 gene. Very little was found in the literature on the functions of the ST6GAL1 gene in pigs, although it has been well studied in other mammals, especially human. The limited information of pig ST6GAL1 gene in the previous pig sequence assembly 10.2 hindered the functional study of ST6GAL1 in a genetic approach.

To better understand the pig ST6GAL1 gene, we first experimentally determined the missing 5' UTR region of the pig ST6GAL1 gene. The transcriptional profiling of the pig ST6GAL gene was then constructed based on the 5' divergent TSSs of different transcripts. 5' alternative splicing pattern of ST6GAL1 transcripts we found is consistent with what was reported in the pig assembly 11.2 released in 2017. Moreover, we found out that the pig ST6GAL1 gene has a strong expression in the liver, which confirmed the claim that ST6GAL1 expression pattern is conserved among species (Meng et al., 2013). Additionally, it clearly demonstrates the use of 5'TSS and the presence of 5' UTEs in different transcripts of tissues. An implication of this is the possibility that multiple promoters might orchestrate the preferential

expressions of transcripts in various tissues. To the best of our knowledge, this is the first study on the identification of the tissue-specific expression pattern of the pig ST6GAL1 gene.

During the study, we repeated the 5' RACE experiments to eliminate the possibilities that the expression of some transcripts was instable. From all three independent biological repeats, we found out in each tissue, some products were efficiently amplified in more than two repeats, indicating they are the dominant transcripts in the corresponding tissue. Then, we grouped 24 mRNA transcripts identified ten mRNA families according to 5' divergent sequences. The transcripts enriched in the respiratory tract—lung, bronchus, and trachea, were selected as the target (FAM 5 and 6).

It has been reported that the region of 20~100 bp flanking 5' TSS is important for efficient initiation (Rach et al., 2009) and elongation of the transcription (Adelman and Lis, 2012). Considering this, we decided to delete ~500 bp upstream and downstream of the two untranslated exons in FAM 5 and 6. The transcriptional profiling of the ST6GAL1 gene was expected to be disrupted due to the inefficient gene transcription in a promoter-regulated fashion. The above deletion of the 5' TSS region yielded a ST6GAL1ΔP genotype (ΔP for short). We found that the resulting model had reduced enzymatic activity of the ST6Gal1 protein, i.e., reduced α 2,6-SA cell surface expression. Furthermore, the results of the viral assessment with low doses and single-cycle infections showed that ΔP cells were moderately resistant to the human IAV strain.

Initially, we expected ΔP cells to have reduced α 2,6-SA expression. Thus, it would be more difficult for the IAV strain with an α 2,6-SA binding preference to

infect ΔP cells. Our viral assessment with Cal04 virus, a human isolated IAV preferentially binding to α 2,6-SA-containing receptor (private communication with John McCauley lab), confirmed our hypothesis: compared to the WT cells, ΔP was less susceptible to Cal04 virus due to a paucity of suitable SA-containing receptor. With Swine87 viral strain, we found out that ΔP cells did not appear to be susceptible to this viral strain according to our experiment. This result is difficult to explain since the binding preference of Swine87 virus was unclear. A possible explanation is that the susceptibility to Swine87 virus is related to the expression levels of both α 2,6-SA and α 2,3-SA. ΔP cells had reduced α 2,6-SA expression, but increased α 2,3-SA expression, so Swine87 virus can still effect ΔP efficiently.

6.2 Alternative methods

6.2.1 Other techniques

To achieve the perturbation of the transcriptional profiling of pig ST6GAL1, we could have alternatively identified the lung-specific enhancers by investigating the histone signatures of the lung-specific enhancer activities. Tissue-specific enhancer elements can be identified by the Chromatin immunoprecipitation with massively parallel sequencing (ChIP-seq) in various tissues. ChIP-seq is a robust and well established method to study the DNA-protein binding pattern (Visel et al., 2009). I met Prof. Zhou's group from the University of California, Davis at ISAG 2017 (Dublin). They have established a ChIP-seq library in pig lung, liver and spleen. Moreover, the sequences derived were aligned to the previous pig assembly 10.2, in which the ST6GAL1 gene was on an unplaced scaffold in the pig genome. The sequencing data showed that the region GL894735.1 (46680-46946) was identified to contain a lung-specific enhancer, rather than a liver or spleen one. This finding is also supported by histone modifications, such as H3K27ac and H3K4me1 markers, which are frequently used for enhancer states (spicuglia and Vanhille, 2012). Therefore, we could delete this lung-specific enhancer region using the CRIPSR/Cas9 system. This would result in edited pig cells or pig models harbouring disrupted transcriptional activity of ST6GAL1 in lung, and thus less human IAV viral recognition and propagation in lung can be achieved. Moreover, this will be unlikely to trigger any excess immunodeficiency since the α 2,6-sialyl-glycans receptors are intact on the surfaces of the liver and the spleen.

Furthermore, cap analysis gene expression (CAGE) tags, which can be used to identify tissue-specific TSSs and promoter usages (Takahashi et al., 2012), could also

be useful. Applying this analysis, the 5' TSS regions of ST6GAL1 transcript variants could be identified in the lung, the bronchus and the trachea, specifically.

If enhancers or other endogenous promoters can be identified specifically regulating airway transcripts of ST6GAL1, we could use co-expression of Cas9 and multiple sgRNAs to precisely excise the epigenome markers and regulatory element regions from ST6GAL1. Alternatively, we could design sgRNAs targeting the regulatory elements of interest. A catalytically-dead Cas9 (dCas9) can be fused to effector domains, increasing or reducing the gene expression by recruiting transcriptional co-activators or co-repressors, respectively (Anton et al., 2019).

To improve the efficiency of mutations in the coding region, the *piggyBac* (PB)-based transposon system could be used. This system is based on an alternative non-viral mobile genetic element. PB system can carry up to 100 kb exogenous DNA, allowing the integration of transposon DNA to 'TTAA' sequences distributed on the genome, and mediating precise genome excision in mammalian cells (Qian et al., 2014). In the future, when attempting to introduce specific mutations in the genome, we could try to combine the CRISPR/Cas9 system and the PB system, which could insert the CRISPR machinery into mammalian genome. The combination of CRISPR and PB system has been successfully applied in the study of human immunodeficiency virus type 1 (HIV-1) inhibition by targeting C-X-C chemokine receptor type 4 (CXCR4) receptor gene for HIV-1 in an HIV-1 reporter cell line, resulting in undetectable exogenous DNA and off-target effects in the genome (Liu et al., 2018).

To better understand the susceptibility of ΔP and ΔCD cells to different IAV strains, future experiments could include the characterisation of receptor binding specificity. We could test binding preferences of viral strains using chicken red blood

cells (CRBCs), which could be modified to exclusively express SA α 2,6-Gal or SA α 2,3-Gal (Glaser et al., 2005). Alternatively, we could perform HA binding assay. Purified HA from IAV isolate of interest can bind to NPTr cells. The alteration of HA binding in edited and WT cells could indicate the altered susceptibility of these cells (Julianna Han et al., 2018)

6.2.2 Other targets

We could also consider decreasing the α 2,3-SA receptor expression on the cell surfaces. α 2,3-SA receptors are catalysed by the ST3Gal1 protein, which is evolutionarily homologous with ST6Gal1 protein, containing identical and conserved sialyl-motifs (Ortiz-Soto and Seibel, 2016). We preliminarily measured its coding gene ST3GAL1 expression levels in pig tissues, similar to the work on ST6GAL1 (**Chapter 3**). The results demonstrate the variant expressions of ST3GAL1 transcripts, indicating its tissue-enriched transcriptional profiling in pig tissues (**Figure 6. 1**). Therefore, we could perform the same procedure for ST3GAL1 manipulation, and identify the airway-specific/abundant regulatory elements. Then, the precise engineering of ST6GAL1 and ST3GAL1 can be performed simultaneously, and the models' susceptibility to different IAV strains could be characterised.

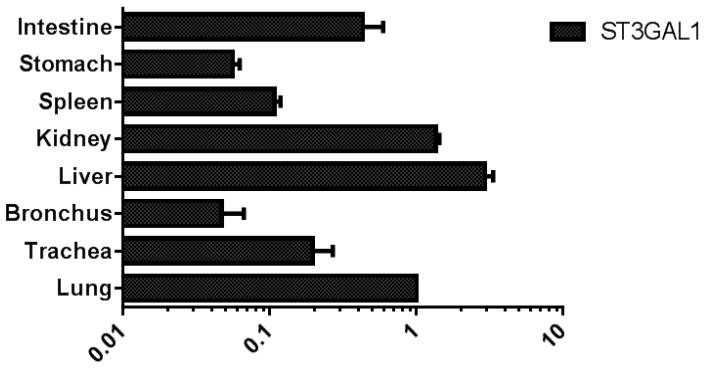


Figure 6. 1 Relative expression of in eight pig tissues.
The same procedure of qPCR was conducted as described in the **Section 2.3.2.3**.
Forward primer: 5'-CGGGAGAAGCAGCCCAATAA-3'
Reverse primer: 5' GCTCTCGGGGTACACGAAAT-3'.
x-axis: relative expression level.

Alternatively, we could study virus-host interactions based on CRISPR/Cas9 genetic analysis, especially the loss-of-function analysis of host factors that are important for IAV invasion. Recently, researchers have established the A549-Genome-scale CRISPR/Cas9 Knockout (A549-GeCKO) library to screen and identify various host factors involved in IAV infections (Julianna Han et al., 2018). Of all candidates identified, Golgi cytidine 5'monophosphate (CMP) sialic acid transporter, SLC35A1, were selected. SLC35A1 can translocate the CMP-sialic acid from cytosol to Golgi lumen, where Sialyltransferase can use it as donor substrate for the synthesis of sialy-glycoconjugates, which are then assembled into host receptors for viral recognition. It has been shown that CHO cells deficient in SLC35A1 are resistant to the Minute virus of mice, and that there is no detrimental impact on cell growth, viability and productivity (Mascarenhas et al., 2017). However, the role of SLC35A1 remains unclear in pigs, and thus it could be a future target. We could attempt to

identify its transcriptional profile in different pig tissues to determine whether it has tissue-associated promoter or enhancers. If so, we could then perform the CRISPR/Cas9 system to precisely engineer the airway-expressed transcripts.

The mechanism of how IAVs infiltrate cells is still not fully understood, especially regarding why hosts are infected in the absence of the virus' preferred SA-containing receptors on the cell surfaces. In particular, the ΔCD cells, having functional knockout of the ST6GAL1 gene, are less susceptible to human IAV thanks to the reduction of the sialyl-glycans receptors on the edited cells, but these cells can still be infected (possibly through other receptors). This is also proved by a study in which SA-removed cells after sialidase treatment can still be infected by IAV efficiently. Therefore, alternative host receptors for the IAV recognition should be explored and elucidated in future, and more factors and host mechanisms are to be studied in-depth.

6.3 Future directions

Following the work described in this study, some follow-up experiments can be performed to further extend the story. The first one is to isolate more pig tissues, so that the transcriptional profiling of pig ST6GAL1 could probably be enriched. As previously described, we isolated nine pig tissues: lung, trachea, bronchus, liver, spleen, kidney, stomach, small intestine, and heart. These tissues broadly cover physiological systems. The reason we did not include other tissues from reproduction, skin and nervous systems, etc., is that these tissues are less associated with immune functions than the selected ones. In the further work, more tissues could be included in the identification of the transcriptional profiling of the ST6GAL1 gene. This helps us improve the understanding of the tissue-expression pattern of ST6GAL1.

We could also include the detection of innate immune responses of IAV infected ΔP and ΔCD cells, such as interferon (IFN) type I and III responses (Kreijtz et al., 2011), and the cytokine signalling (Delgado-Ortega et al., 2014).

Since the *in vitro* model is not fully representative of the whole pathophysiology, the *in vitro* strategy evaluated in this study can be applied to generate an *in vivo* pig model. If SA α 2,6-Gal receptors for IAV entry could be eliminated in the airway exclusively, the pigs could be protected from the infection by IAV with α 2,6-SA preference, thanks to the lack of SA α 2,6-Gal receptors. Moreover, the edited pigs are entitled to retain normal biological functions conducted by α 2,6-sialylated glycan receptors. Therefore, the cross-transmission will be reduced in pig herds and human populations.

Although it is anticipated that the edited pigs might not be resistant to avian IAV infection, they are likely exhibit a partial elimination of human IAV replication,

which will be an important improvement over the current situation. The viral reassortant events would also be limited to some extent, and this is beneficial to animal welfare, human public health, and economy.

The best method to create a partial ablation of SA α 2,6-Gal receptors exclusively in the airway of pigs is the CRISPR/Cas9 system, which enables scientists to modify the genome with unprecedented simplicity (Cho et al., 2013). However, when performing an *in vivo* experiment, there are still some open questions: (1) how to validate and enhance the on-target efficiency, (2) how to reduce off-target possibilities, (3) and how to reduce the cytotoxicity associated with the CRISPR/Cas9 system.

The first question is how to increase the on-target efficiency. PAM variance is a critical factor (Kleinstiver et al. 2015). Apart from the canonical PAM ('NGG') in the *Sp*Cas9 system, 'NAG' and 'NGA' are also associated with cutting activities (Zhang et al. 2014). Additionally, we could increase the specificity by activating *Sp*Cas9 with some factors, such as some small exogenous molecules. Moreover, new variants of Cas9 have been created that have higher fidelity, such as the HF *Sp*Cas9 (Kleinstiver et al. 2015) and e*Sp*Cas9 (Slaymaker et al. 2016).

The second question is how to reduce the possibility of off-target effects, which is the most critical issue for genome editing. To minimise the off-target effect, various approaches have been used (Tycko et al., 2016). For instance, sgRNA truncation (Fu et al. 2014), 5' extension (Kim et al., 2015), the use of the HF *Sp*Cas9, and the transient delivery of ribonucleoprotein complex (RNP) involving gRNA and Cas9 by cationic lipids (Zuris et al., 2014).

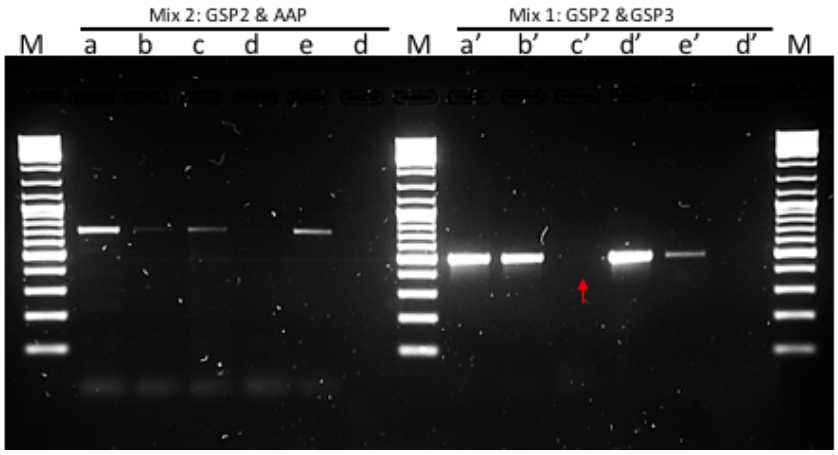
The third question is how to reduce the cytotoxicity and to increase embryonic survival rate. To this end, we could directly inject into embryos Cas9 mRNA and sgRNAs, which have been shown to be low in cytotoxicity during the embryo development (Hai T et al., 2014). Alternatively, we could inject into embryos a low concentration of CRISPR/Cas9 RNA. But the possibility of mosaic genotypes caused by injecting CRISPR/Cas9 system can lead to unexpected phenotypes of the founder pigs. To reduce the bad effects, other Cas orthologous have been developed (Ruan et al. 2017) incorporating CRISPR recognition sites into the *Sp*Cas9 cassette to shorten Cas9 protein expression period. In future, we could inject CRISPR reagents into the cytoplasm of one-cell zygote-stage pig embryos and generate gene modified pigs.

To make the story more complete, IAV challenge will be included in edited pigs. After viral infection the *in vivo* pig models, the significance of ST6GAL1 for host homeostasis, and immunological pathways should be investigated in depth. Moreover, ethical and legal issues should be considered.

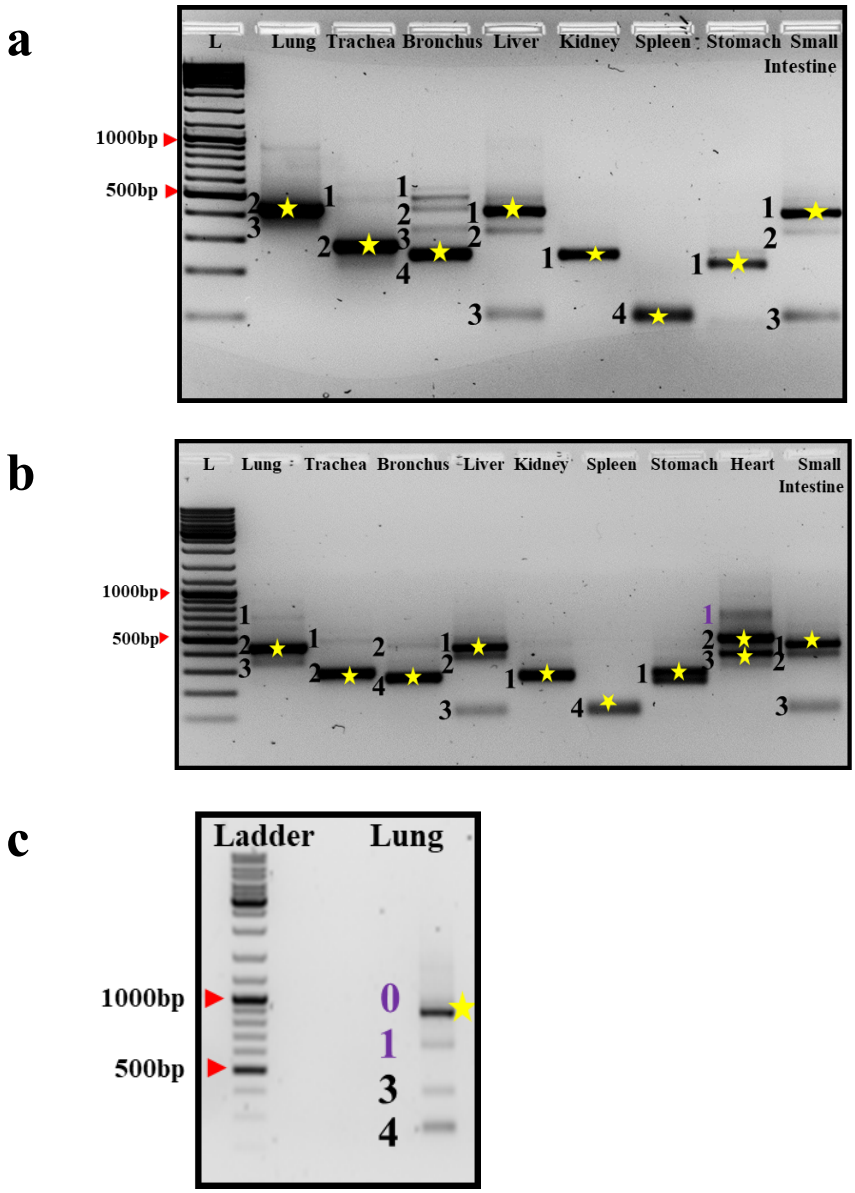
6.4 Conclusion

The experiments precisely engineering pig ST6GAL1 gene suggest that the gene is essential for the biosynthesis of SA α 2,6-Gal on cell surfaces. Moreover, the manipulation of the ST6GAL1 gene renders the edited trachea cells less susceptible to the human IAV. Overall, our work presented the first study on host IAV receptor manipulation in pig cells, and which helps deepen the understanding of the mechanisms involved in virus-host interactions; it serves as a primitive but valid strategy for modelling host resistance against IAVs; moreover, once applied, our 'intervening approach' has the potential of preventing future pandemics in humans. Furthermore, host resistance against more IAV strains can be modelled in a similar fashion.

Appendix



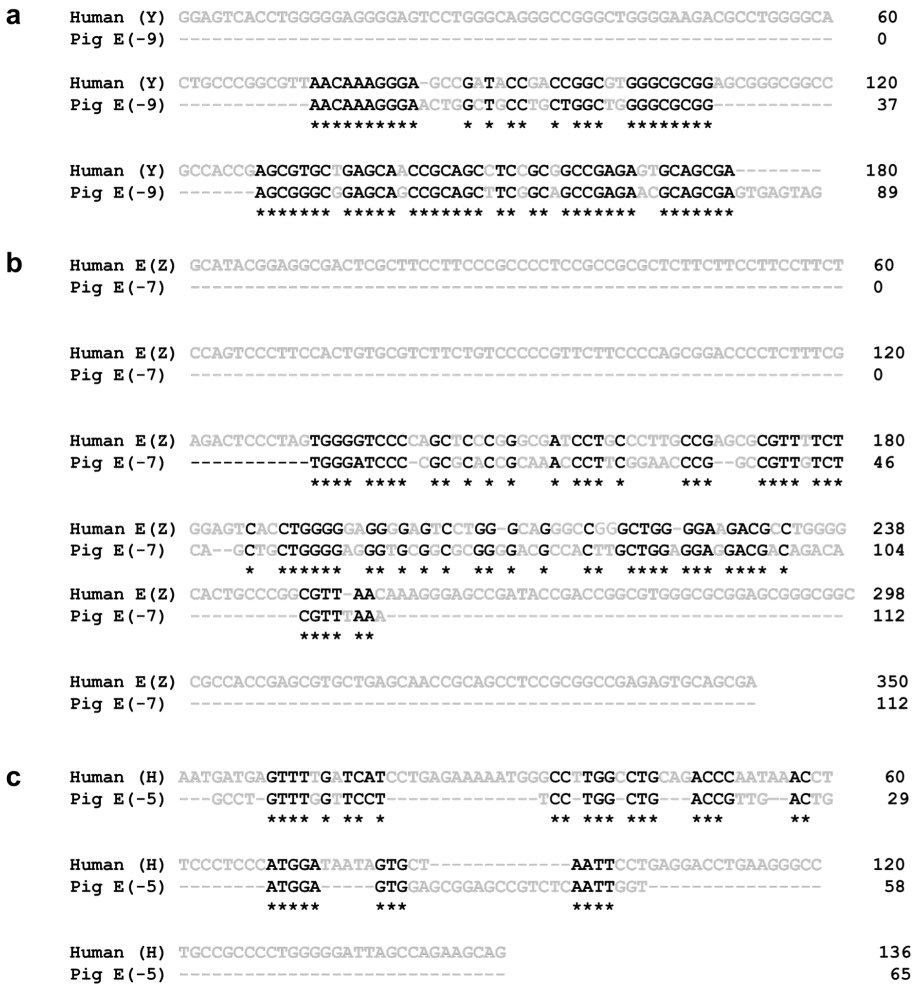
Appendix 1: Positive control of 5' RACE. The control cDNA, 100 µg of an *in vitro* transcribed RNA from the chloramphenicol acetyltransferase (CAT) gene, has been engineered to contain a 3'poly (A) tail.
The RNA template was then degraded using RNase Mix. DNA was amplified using the 5' RACE Anchor Primer and GSP2 (provided). Mix1 was amplified using GSP2 and GSP3, and Mix2 amplified using GSP2 and AAP. a and a' were cDNA products; b and b' were purified cDNA; c and c' are tailed cDNA after TdT tailing reaction (PCR), indicating the successful tailing procedure; d and d' were non-tailed cDNA products; and e and e' were water control DNA. The c' product (red arrow) was very faint, indicating the low efficiency of tailing procedure in Mix1.



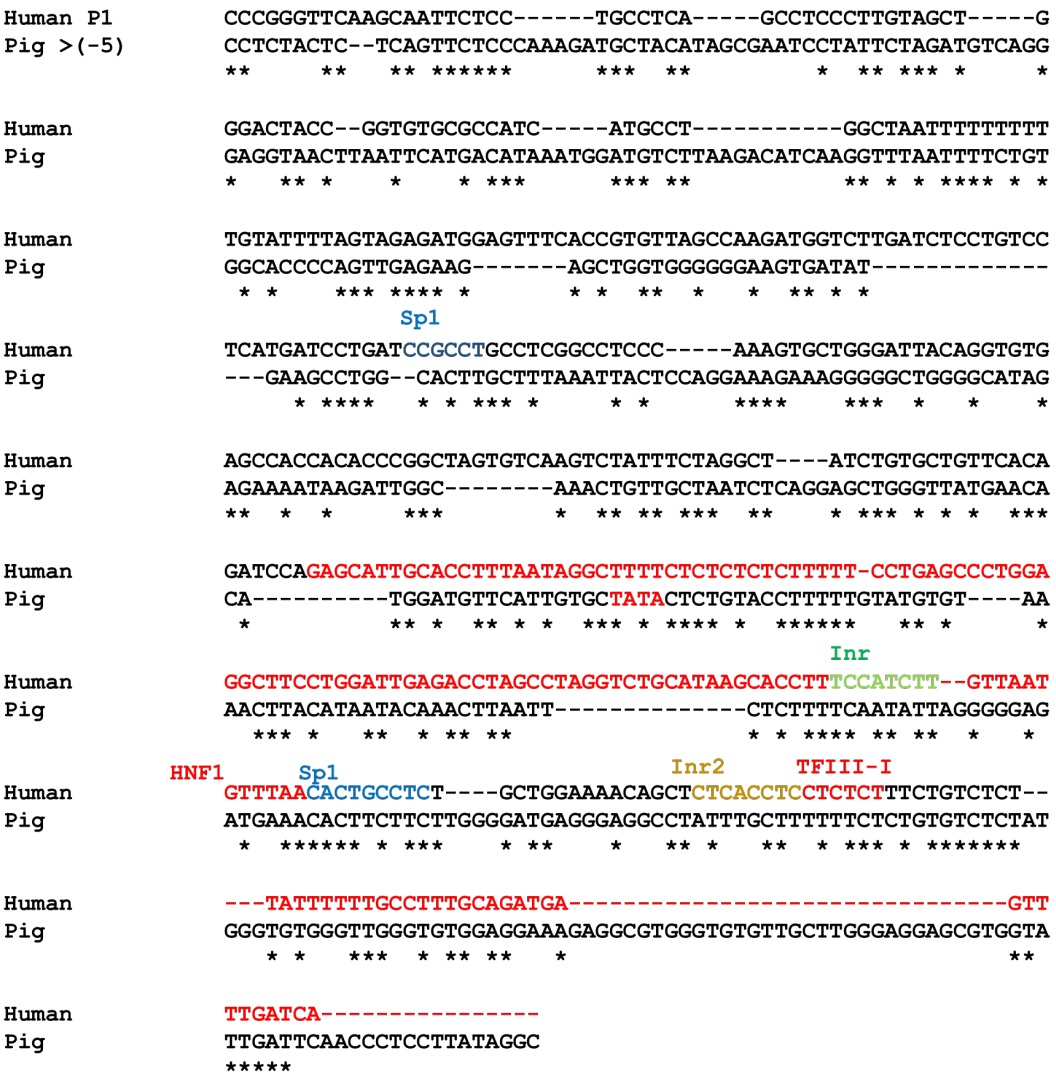
Appendix 2: 5' RACE PCR results. All numbered products were extracted and corresponding with each other in each experimental repeat. Total 24 transcripts were obtained, and the highest expressed transcripts in each tissue were pointed by yellow asterisks on the 2% agarose gel. The lung was repeated for the fourth time (c). However, we just amplify lung1 for once and sequencing of this sample was failed, indicating it might be a non-specific band or other reasonable reasons. Also, we did not include heart in the third repeat (b), because the PCR using heart cDNA was efficient in the first and second results. L = ladder.

```
E(-9)  5′-AACAAAGGGA ACTGGCTGCC TGCTGGCTGG GGCGCGGAGC GGGCGGAGCA GCCGCAGCTT CGGCAGCCGA
           GAACGCAGCG AGTGAGTAG

E(-8)  5′-TGGGATCCCC GCGCACCGCA AACCCTTCGG AACCCGGCCG TTGTCTCAGC TGCTGGGGAG GGTGCGGCGC
           GGGGACGCCA CTTGCTGGAG GAGGACGACA GACACGTTTA AA

E(-7)  5′-AGGCAGAGTA GAGTGAATTG TGCAGAGACC TGCAGAGAGC AGCCCAGAGA GCAGCAGTGG CGTCTTGGAC
           CACAGCTGAA ATGAATTGGC CTACTTGAGC CCGCCCAAGC CCTGGG

E(-6)  5′-GTTCGGCTTT ACTGGGCACC ATTTATTAAG CATCTGCCAT GTTCCAGGCA TGGCAGGATC CCTCCTTCTA
           GCCCTTCCAG CAGCCCTGTG AGG

E(-5)  5′-GCCTGTTTGG TTCCTTCCTG GCTGACCGTT GACTGATGGA GTGGAGCGGA GCCGTCTCAA TTGGT

E(-4)  5′-GTGTGGTCAT CAAACGTGTT TTGAGTGCAG ACTGCAGGCT GGGTAGAGGT GGAAGAGACC GCAAGGTGCC
           TGCACCCACT AGAGAGCAAG TGTCACAGAA CAAAGTACAG AACCTGGGGC GCTTCGGAGT ATGGAGGTGA
           GGAAGATGAG GAGGAACCAG CATTAGAGAT TATTGAGGAG GAGCACCCGG GAAGGTGGAG AAAGACCCCC
           AAACT

E(-3)  5′-ACTTATTTTT GATCATTCTG AGAGAAGTGA GTCTTGGCCA CCAGACCCAA TAAACCAAAC ATCCCTCCCA
           TGGAGAATTG TGCTGGTTCT TGAGGACCTG GAGGGCCTGC CGCCCCAGGG GAATGAGCCA GAAGCAGG

E(-2)  5′-ATGCCGGTCA AAGAAGGACT GGTTGGCACC CAGGTCATAG TCGAAGTGTG TGGCCTCCTC GGGCTTCCTC
           CAGCCAGCCA GCCACGACCC CAGCAGAGTG CGCAGCATTG AGTGG

E(-1)  5′-GGCCTGCTTT TCTGCTCAGG ACAGAGTGGT TTCCTTGAAC ACATCTTTCA TT

                +1
E(1)   5′-ATGATTCAC ACCAGCTTG AAGAAAAAA TTCAGCTGC TGCCTGCTG GTCTTTCTC CTGTTTGCG
           ATCATTTGT GTGTGGAAG GAAAAGAAG AAAGGGAGT TACTATGAA TCCCTTAAA TTGCAAACC
           AAGGAGCTC CAGGTGCTG AGGAGTCTG GAGAAACTG ACTCCAGGG TCTGGCTCC CAGTCTGCA
           TCCTCCAGC GGCAC GTGCCGCTGGAGGATGCAGA--3′
                          Nested primer
```

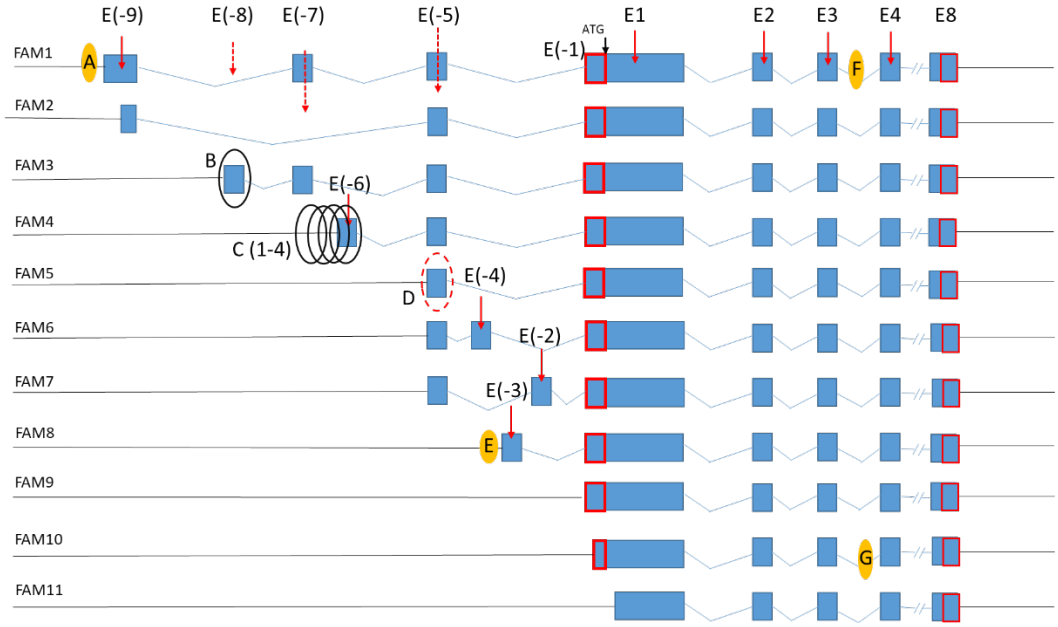
Appendix 3: 5' UTR sequencing results. (a) Exons were displayed from E(1) through E(-9), Nested primer was located in E(1).



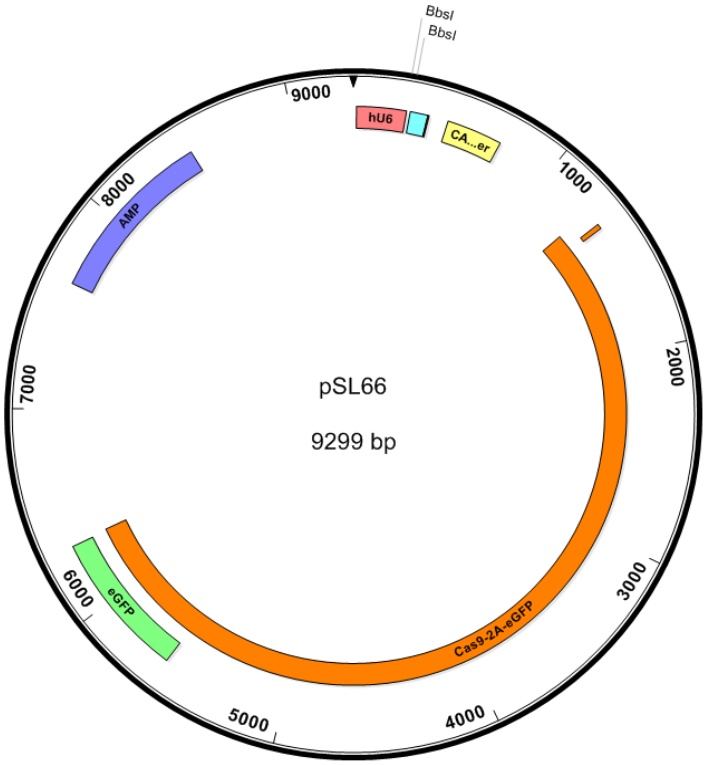
Appendix 4: Sequencing alignment of 5'UTR. Alignment of ST6GAL1 untranslated E(H), E(Z) and E(Y), and pig ST6GAL1 untranslated exon E(-5), E(-7), and E(-9). The asterisks showed DNA identity. The data was aligned by Clustal Omega. Accession number of E(X) was X62822, and of human E(Y) and E(Z) was X17147.



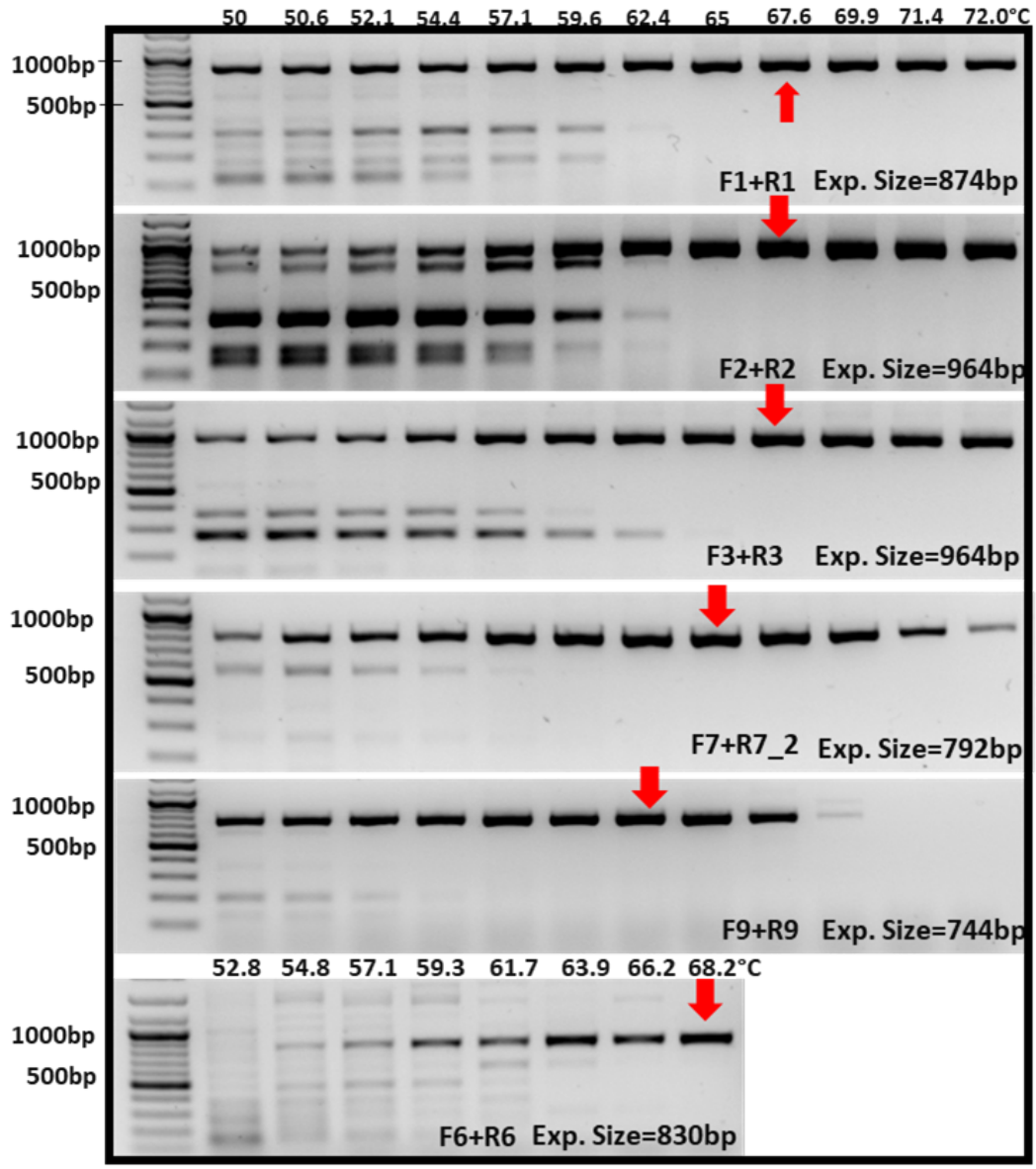
Appendix 5: Sequence alignment of P1 region. 500 bp upstream of E(-5) was aligned with human P1 region (LOC111242783). The asterisks showed DNA identity. The nucleotides written in red, green, and were the promoter elements in human P1 region. TCCATCTT (green): initiator element, as a core promoter, is important in the transcription initiation. Sp1: specificity protein 1, a transcription factor. HNF1: homeobox promoter. TFIII-I: transcription factor. (Lorena et al., 2012). The data was aligned Clustal Omega. Gaps were indicated by the line.



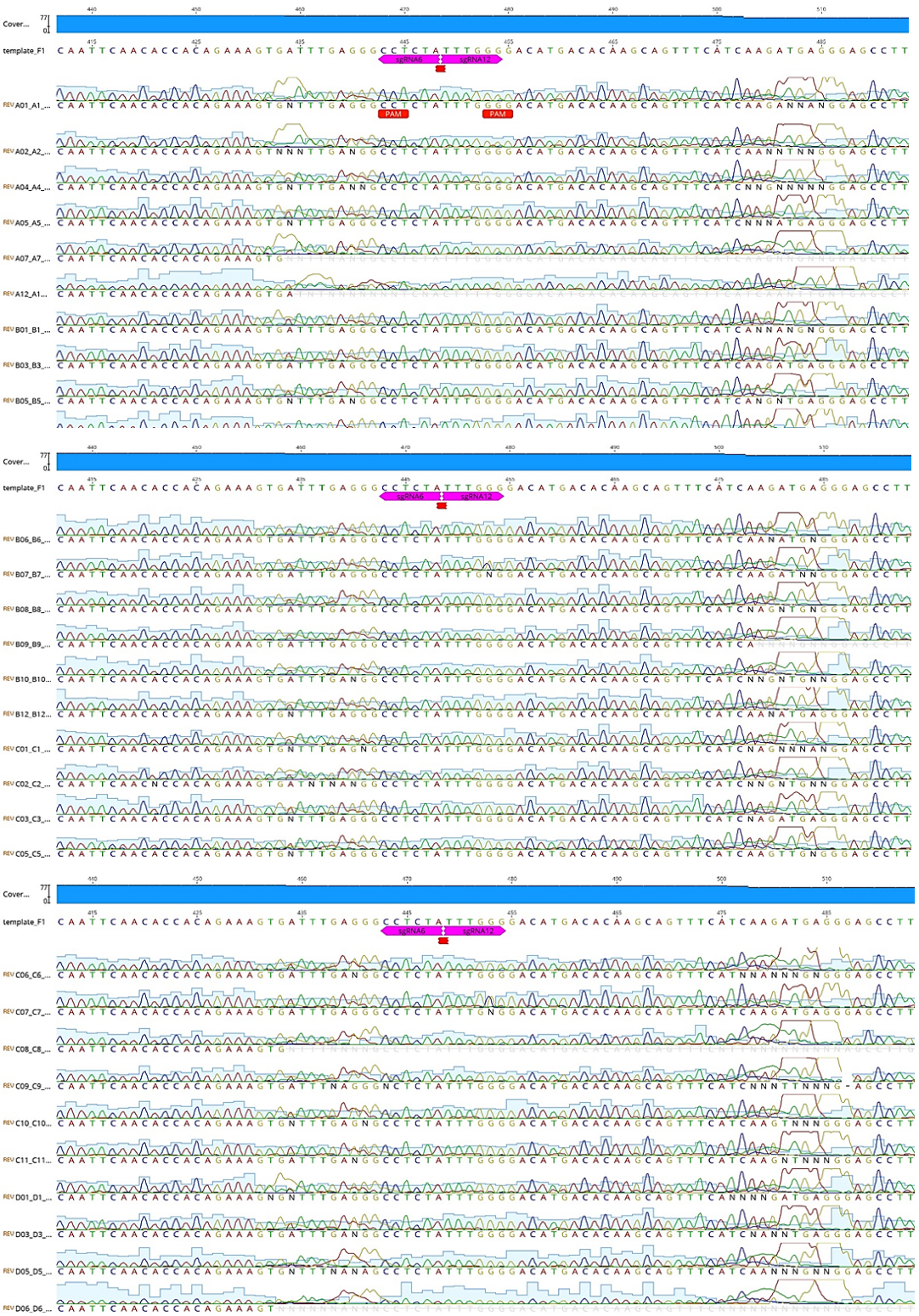
Appendix 6: Prediction of regulatory features.
There are seven predicated promoter motifs predicated and labelled in a circular symbol in this schematic map, ranging from A to G. C(1-4) represented four predicated elements collectively or individually. This image was originally constructed by Sylvia Beka. The software and tools used in this analysis were described in **Section 2.3.1.8**.

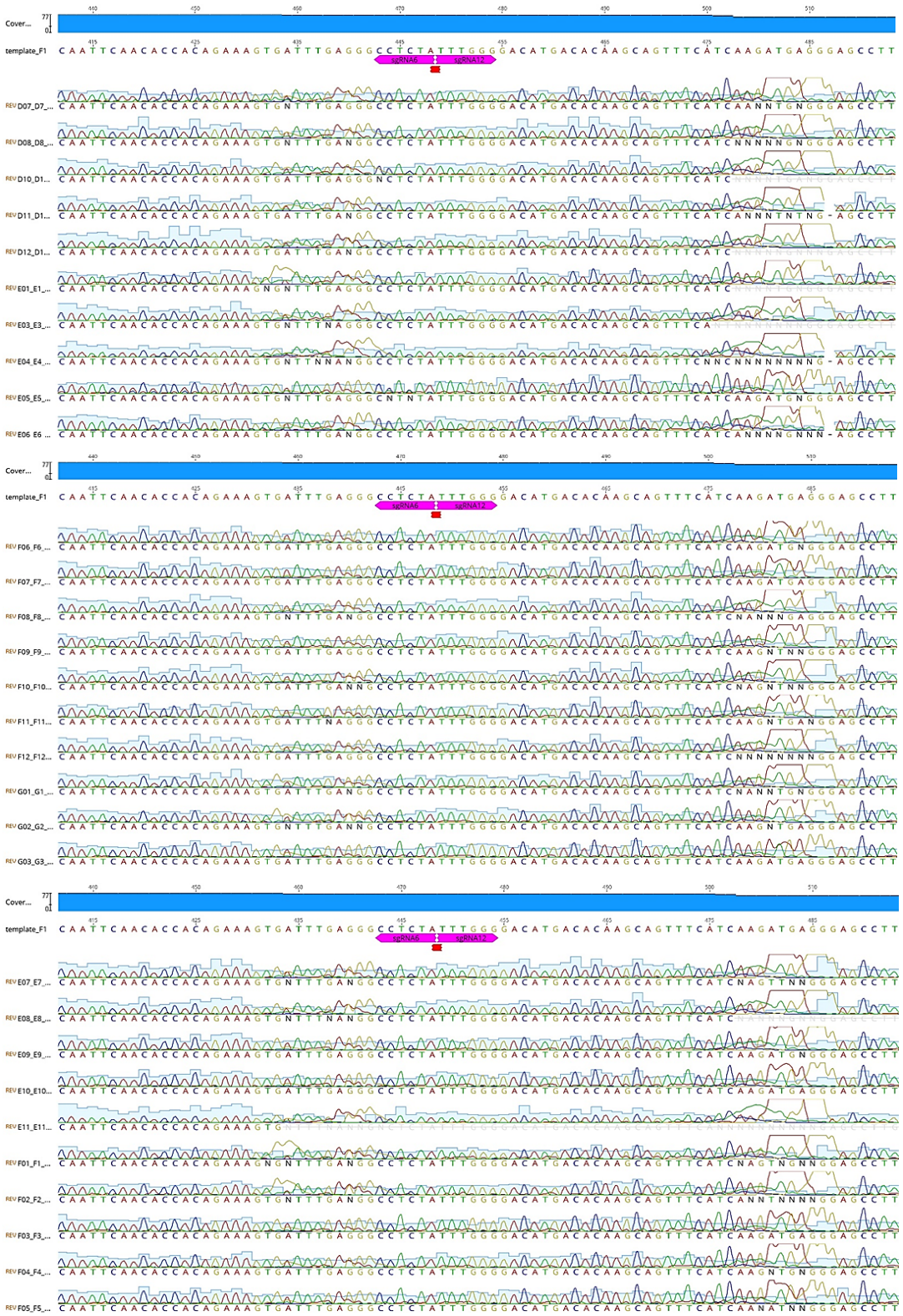


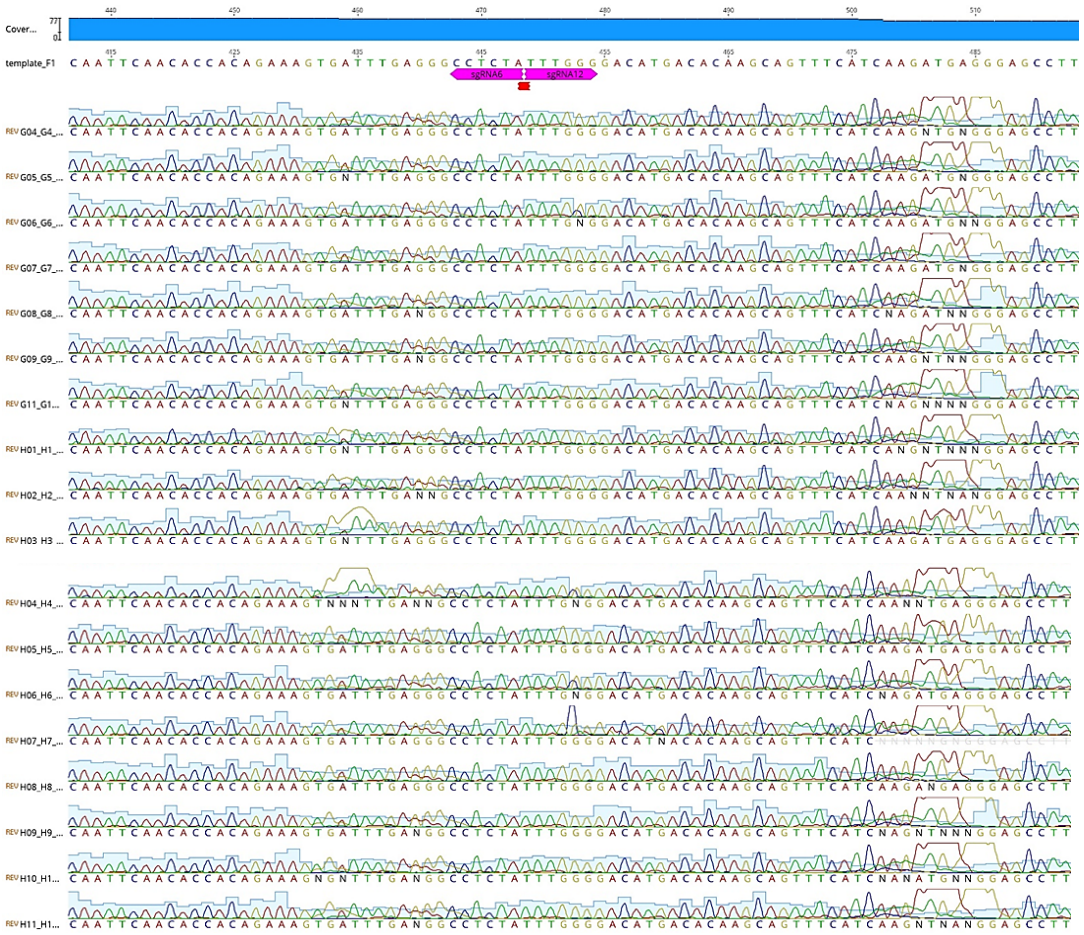
Appendix 7: The map of the pSL66 plasmid. It was used to encode for the CRISPR-Cas9 complex. The plasmid contains an ampicillin gene (AMP, dark blue) to aid colony selection. An fluorescent tag (eGFP, green) on the C-Termini of Cas9-2A protein (orange) was used to identify the transfected cells. A hU6 promoter (red) was upstream of the gRNA scaffold (blue) to initiate the transcription.



Appendix 8: Gradient PCR. PCRs were performed across an annealing temperature gradient from 50 ºC to 72 ºC using primer pairs F1+R1, F2+R2, F3+R3, F7+R7_2, F9+R9 and F6 +R6, followed by the visualisation on a 2% agarose gel. They were all primers used in this experiment.

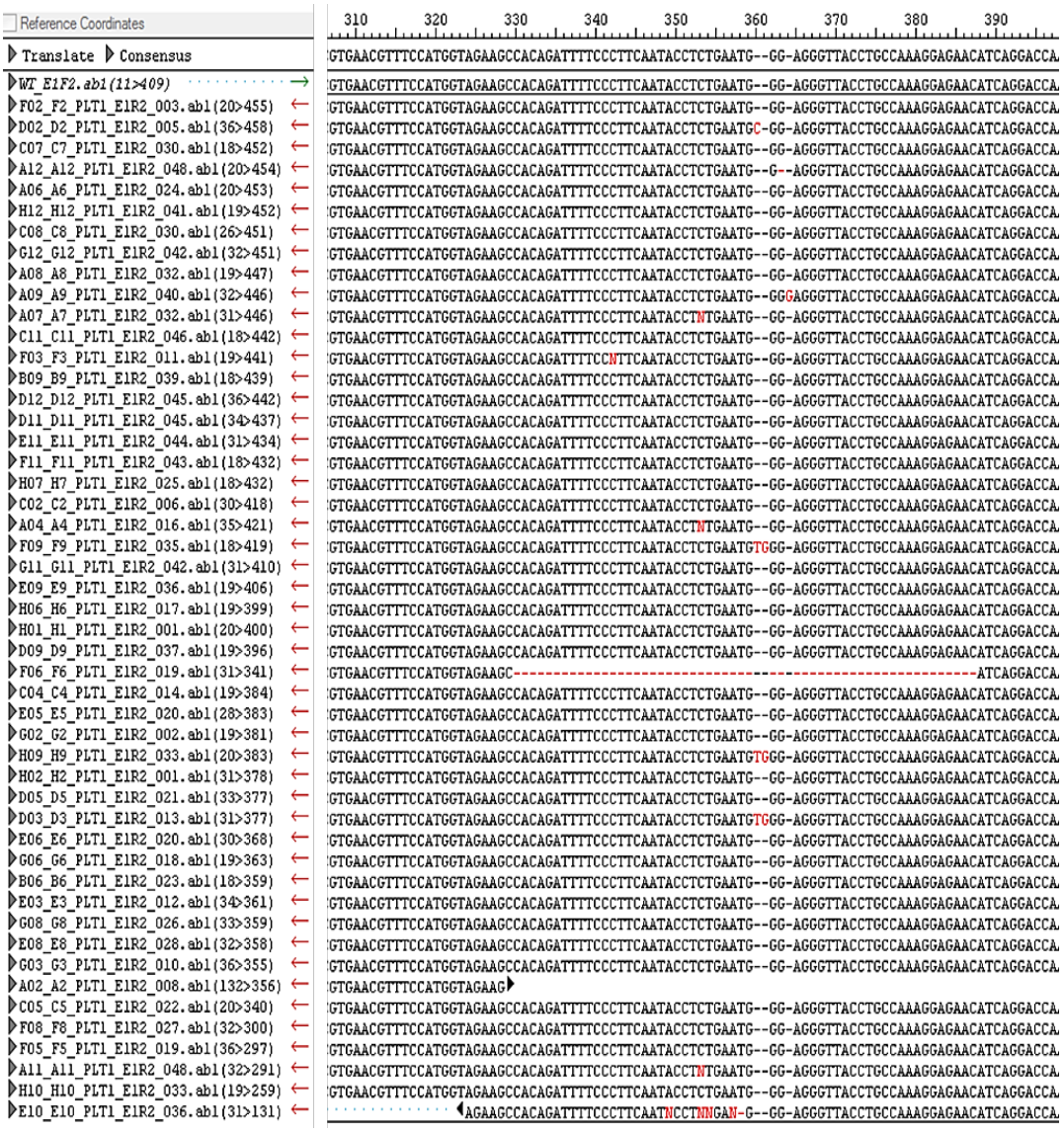

Cover...
template_F1
sgRNA6
sgRNA12
PAM
PAM
REV A01_A1_...
REV A02_A2_...
REV A04_A4_...
REV A05_A5_...
REV A07_A7_...
REV A12_A1...
REV B01_B1_...
REV B03_B3_...
REV B05_B5_...
Cover...
template_F1
sgRNA6
sgRNA12
REV B06_B6_...
REV B07_B7_...
REV B08_B8_...
REV B09_B9_...
REV B10_B10...
REV B12_B12...
REV C01_C1_...
REV C02_C2_...
REV C03_C3_...
REV C05_C5_...
Cover...
template_F1
sgRNA6
sgRNA12
REV C06_C6_...
REV C07_C7_...
REV C08_C8_...
REV C09_C9_...
REV C10_C10...
REV C11_C11...
REV D01_D1_...
REV D03_D3_...
REV D05_D5_...
REV D06_D6_...

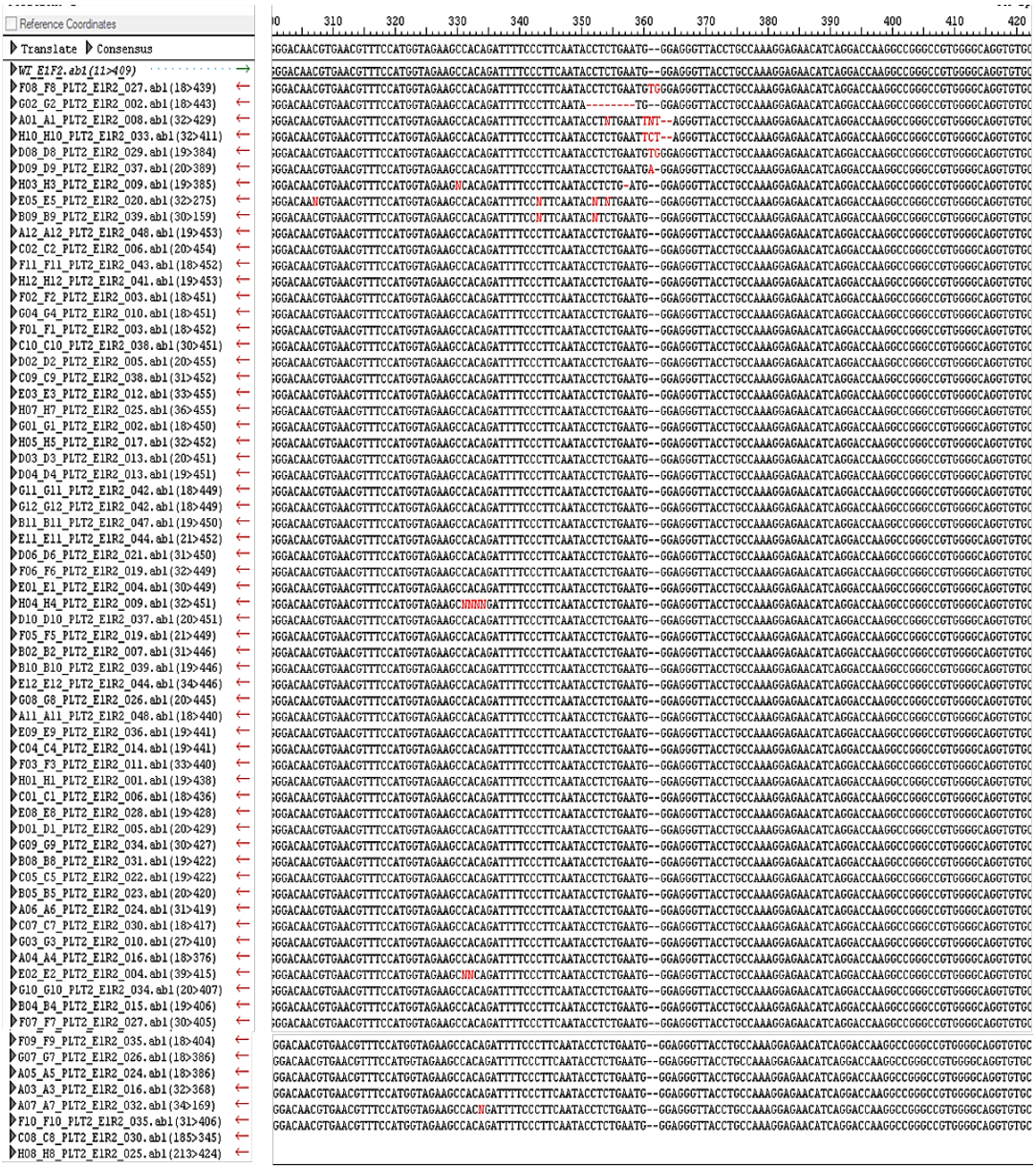

Cover...
template_F1
sgRNA6
sgRNA12
REV D07_D7_...
REV D08_D8_...
REV D10_D1...
REV D11_D1...
REV D12_D1...
REV E01_E1_...
REV E03_E3_...
REV E04_E4_...
REV E05_E5_...
REV E06_E6_...
REV F06_F6_...
REV F07_F7_...
REV F08_F8_...
REV F09_F9_...
REV F10_F10...
REV F11_F11...
REV F12_F12...
REV G01_G1_...
REV G02_G2_...
REV G03_G3_...
REV E07_E7_...
REV E08_E8_...
REV E09_E9_...
REV E10_E10...
REV E11_E11...
REV F01_F1_...
REV F02_F2_...
REV F03_F3_...
REV F04_F4_...
REV F05_F5_...



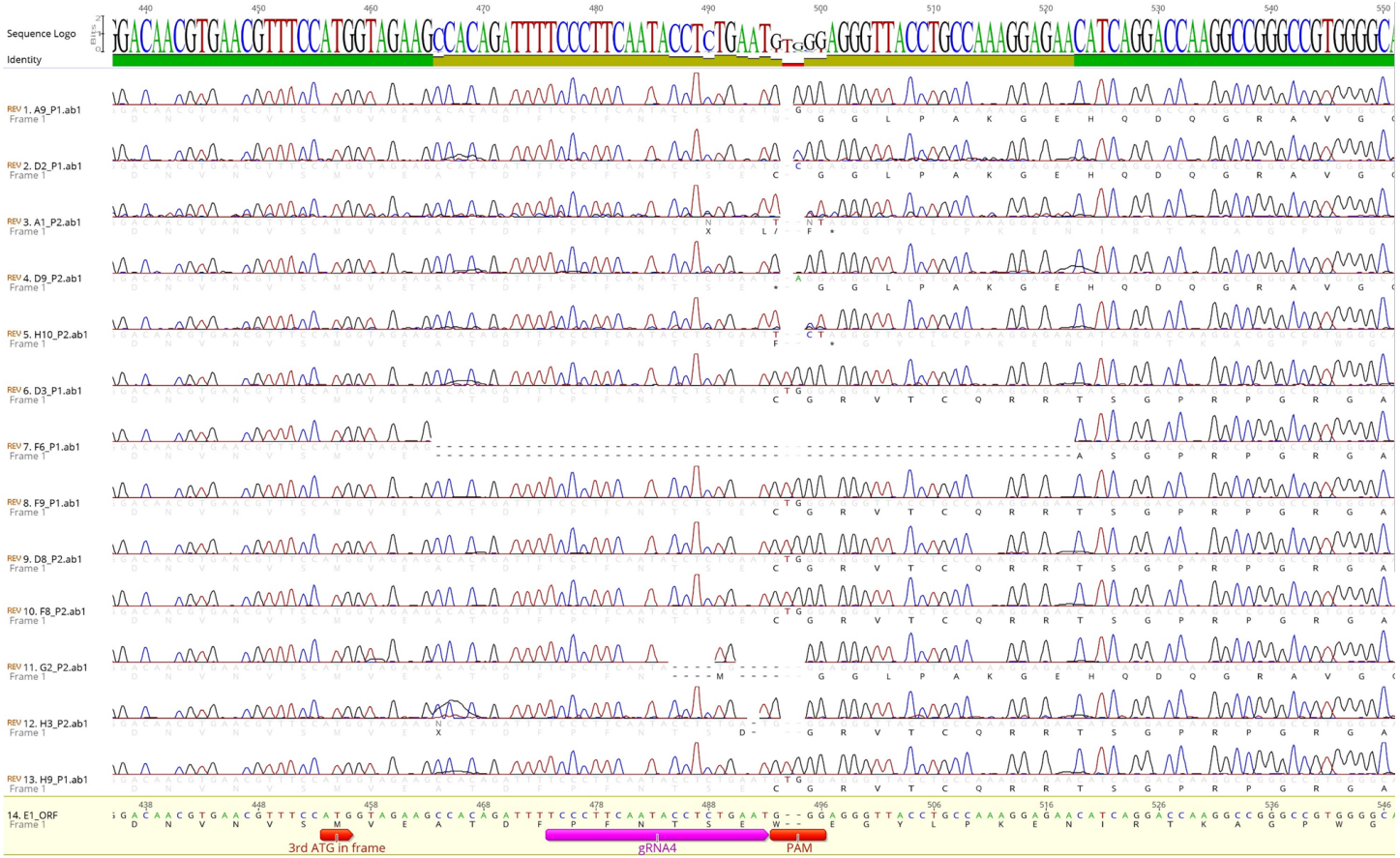
bad samples	good samples						
Orginal location	Seq name (IGMM)	Orginal location	Seq name (IGMM)	Orginal location	Seq name (IGMM)	Orginal location	Seq name (IGMM)
A1	A1	C1	C1	E1	d2	G1	G1
A2	A2	C2	C2	E2	E2	G2	g2
A3	A3	C3	C3	E3	E3	G3	G3
A4	A4	C4	C4	E4	E4	G4	G4
A5	A5	C5	g1	E5	e2	G5	G5
A6	A6	C6	C6	E6	E6	G6	G7
A7	A7	C7	h1	E7	E7	G7	G8
A8	A8	C8	C8	E8	E8	G8	G9
A9	a1	C9	C9	E9	E9	G9	a3
A10	A10	C10	a2	E10	f2	G10	mixed
A11	A11	C11	C11	E11	E11	G11	G11
A12	A12	C12	C12	E12	E12	G12	G12
B1	B1	D1	D1	F1	F1	H1	H1
B2	B2	D2	D2	F2	F2	H2	H2
B3	b1	D3	D3	F3	F3	H3	H3
B4	B4	D4	D4	F4	F4	H4	H4
B5	B5	D5	b2	F5	F5	H5	H5
B6	c1	D6	D6	F6	F6	H6	H6
B7	d1	D7	D7	F7	F7	H7	H7
B8	e1	D8	D8	F8	F8	H8	H8
B9	B9	D9	D9	F9	F9	H9	H9
B10	B10	D10	D10	F10	F10	H10	H10
B11	B11	D11	D11	F11	F11	H11	H11
B12	f1	D12	c2	F12	F12	H12	H12

Appendix 9: Sequencing results of ΔP candidates. Given that the poor growth of some cell clones, we combined all the PCR products from the well-grown cell clones in one 96-well plate for sequencing. The samples grown bad were written in pink colour. For example, B3 in the trace file was actually the result from b1 sample.

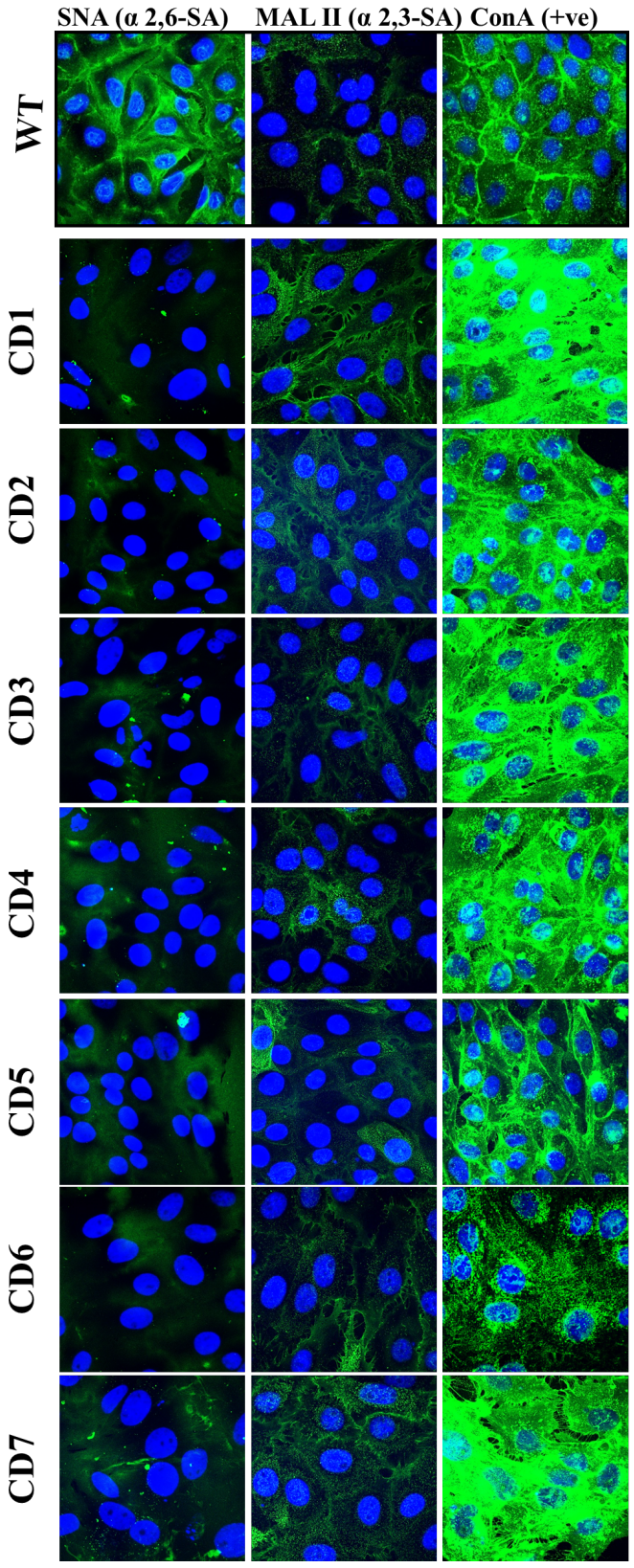

Reference Coordinates
Translate
Consensus
310
320
330
340
350
360
370
380
390
WT_E1F2.ab1(11>409)
F02_F2_PLT1_E1R2_003.ab1(20>455)
D02_D2_PLT1_E1R2_005.ab1(36>458)
C07_C7_PLT1_E1R2_030.ab1(18>452)
A12_A12_PLT1_E1R2_048.ab1(20>454)
A06_A6_PLT1_E1R2_024.ab1(20>453)
H12_H12_PLT1_E1R2_041.ab1(19>452)
C08_C8_PLT1_E1R2_030.ab1(26>451)
G12_G12_PLT1_E1R2_042.ab1(32>451)
A08_A8_PLT1_E1R2_032.ab1(19>447)
A09_A9_PLT1_E1R2_040.ab1(32>446)
A07_A7_PLT1_E1R2_032.ab1(31>446)
C11_C11_PLT1_E1R2_046.ab1(18>442)
F03_F3_PLT1_E1R2_011.ab1(19>441)
B09_B9_PLT1_E1R2_039.ab1(18>439)
D12_D12_PLT1_E1R2_045.ab1(36>442)
D11_D11_PLT1_E1R2_045.ab1(34>437)
E11_E11_PLT1_E1R2_044.ab1(31>434)
F11_F11_PLT1_E1R2_043.ab1(18>432)
H07_H7_PLT1_E1R2_025.ab1(18>432)
C02_C2_PLT1_E1R2_006.ab1(30>418)
A04_A4_PLT1_E1R2_016.ab1(35>421)
F09_F9_PLT1_E1R2_035.ab1(18>419)
G11_G11_PLT1_E1R2_042.ab1(31>410)
E09_E9_PLT1_E1R2_036.ab1(19>406)
H06_H6_PLT1_E1R2_017.ab1(19>399)
H01_H1_PLT1_E1R2_001.ab1(20>400)
D09_D9_PLT1_E1R2_037.ab1(19>396)
F06_F6_PLT1_E1R2_019.ab1(31>341)
C04_C4_PLT1_E1R2_014.ab1(19>384)
E05_E5_PLT1_E1R2_020.ab1(28>383)
G02_G2_PLT1_E1R2_002.ab1(19>381)
H09_H9_PLT1_E1R2_033.ab1(20>383)
H02_H2_PLT1_E1R2_001.ab1(31>378)
D05_D5_PLT1_E1R2_021.ab1(33>377)
D03_D3_PLT1_E1R2_013.ab1(31>377)
E06_E6_PLT1_E1R2_020.ab1(30>368)
G06_G6_PLT1_E1R2_018.ab1(19>363)
B06_B6_PLT1_E1R2_023.ab1(18>359)
E03_E3_PLT1_E1R2_012.ab1(34>361)
G08_G8_PLT1_E1R2_026.ab1(33>359)
E08_E8_PLT1_E1R2_028.ab1(32>358)
G03_G3_PLT1_E1R2_010.ab1(36>355)
A02_A2_PLT1_E1R2_008.ab1(132>356)
C05_C5_PLT1_E1R2_022.ab1(20>340)
F08_F8_PLT1_E1R2_027.ab1(32>300)
F05_F5_PLT1_E1R2_019.ab1(36>297)
A11_A11_PLT1_E1R2_048.ab1(32>291)
H10_H10_PLT1_E1R2_033.ab1(19>259)
E10_E10_PLT1_E1R2_036.ab1(31>131)

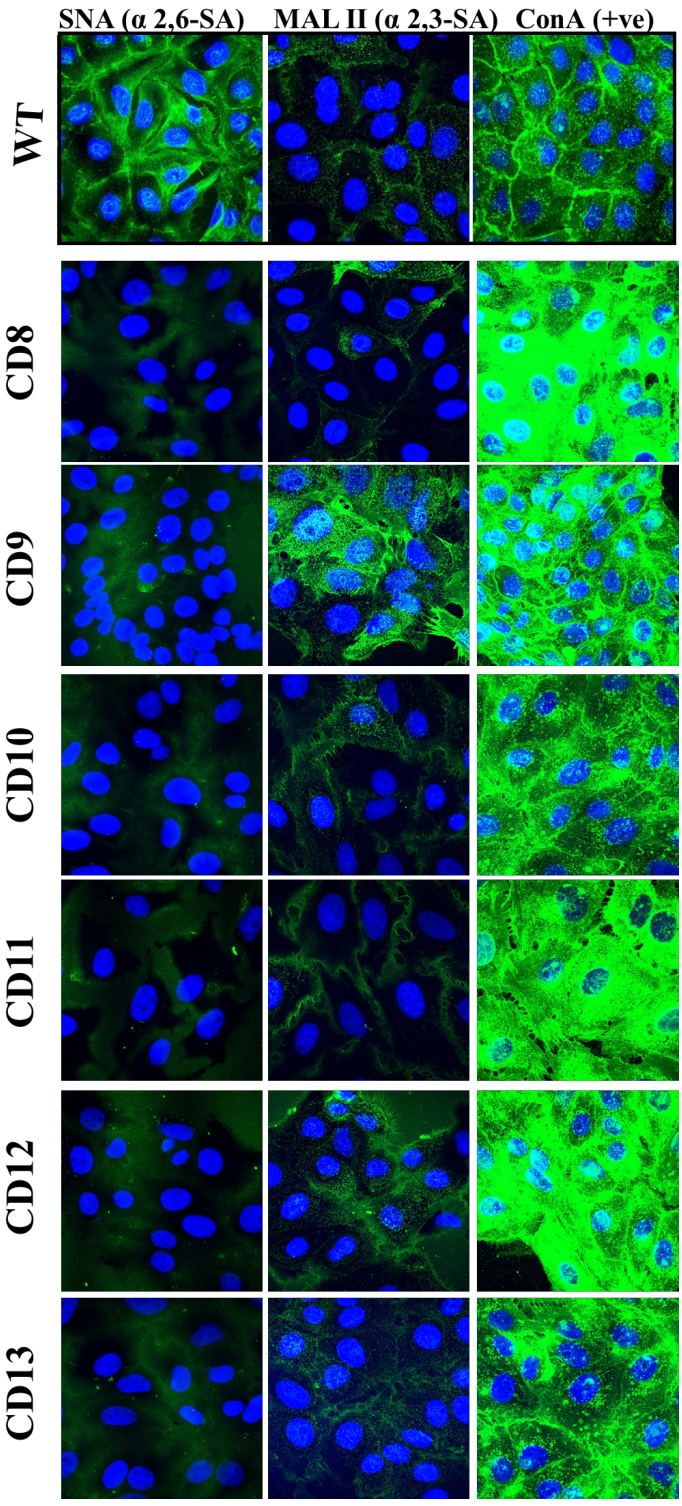


Appendix 10: Sequencing results of ΔCD candidates.

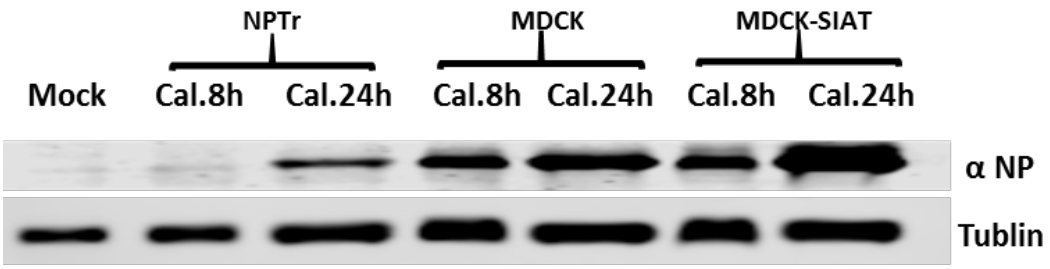


Appendix 11: Chromatogram of genotyping results of ΔCD NPTr cells. This is the supplementary material of **Section 4.4.1.4.** The bottom is the WT control and others are cells carrying KO sequence. Sequencing result indicates that the stop codon introduced are denoted by asterisk.


SNA (α 2,6-SA)
MAL II (α 2,3-SA)
ConA (+ve)
WT
CD1
CD2
CD3
CD4
CD5
CD6
CD7



Appendix 12: The expression levels of SA in all ΔCD cells. This is the supplementary material of **Section 5.2.2.4**. Magnification factor 64x.



Appendix 13: NP expression at 8 and 24 h.p.i with Cal04 virus. As shown here, NP was undetectable at 8 h.p.i so a longer incubation was required (**Section 5.2.3.1.2**). MDCK and MDCK-SIAT1 cells were used here as a positive control. MDCK-SIAT1 presented higher NP production than that in MDCK at 24 h.p.i with Cal04 virus, indicating MDCK-SIAT1 cells (ST6GAL1 overexpression), was more susceptible to Cal04 viral infection.

Reference

AASHEIM, H. C., AAS-ENG, D. A., DEGGERDAL, A., BLOMHOFF, H. K., FUNDERUD, S. & SMELAND, E. B. 1993. Cell-specific expression of human beta-galactoside alpha 2,6-sialyltransferase transcripts differing in the 5' untranslated region. *Eur J Biochem,* 213**,** 467-75.

AB, R., JS, H., RL, P., T, D., DE, S., AM, R., UNITED STATES GEOLOGICAL SURVEY ALASKA SCIENCE CENTER, A., ALASKA, UNITED STATES OF AMERICA., REEVES, A. B., HALL, J. S., POULSON, R. L., DONNELLY, T., STALLKNECHT, D. E. & RAMEY, A. M. 2018. Influenza A virus recovery, diversity, and intercontinental exchange: A multi-year assessment of wild bird sampling at Izembek National Wildlife Refuge, Alaska.

ABCAM. *Counting cells using a hemocytometer | Abcam* [Online]. Available: https://www.abcam.com/protocols/counting-cells-using-a-haemocytometer [Accessed].

ADELMAN, K. & LIS, J. T. 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet,* 13**,** 720-31.

AGRAWAL, A. & SCHATZ, D. G. 1997. RAG1 and RAG2 form a stable postcleavage synaptic complex with DNA containing signal ends in V(D)J recombination. *Cell,* 89**,** 43-53.

AIRD, E. J., LOVENDAHL, K. N., MARTIN, A. S., HARRIS, R. S. & GORDON, W. R. 2018. Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Communications Biology,* 1**,** 54.

ALEXANDER, D. J. 2006. Avian influenza viruses and human health. *Dev Biol (Basel),* 124**,** 77-84.

ANTON, T., DEPARTMENT OF BIOLOGY II AND CENTER FOR INTEGRATED PROTEIN SCIENCE MUNICH (CIPSM), L. M., 82152 MARTINSRIED, GERMANY, KARG, E., DEPARTMENT OF BIOLOGY II AND CENTER FOR INTEGRATED PROTEIN SCIENCE MUNICH (CIPSM), L. M., 82152 MARTINSRIED, GERMANY, BULTMANN, S. & DEPARTMENT OF BIOLOGY II AND CENTER FOR INTEGRATED PROTEIN SCIENCE MUNICH (CIPSM), L. M., 82152 MARTINSRIED, GERMANY 2019. Applications of the CRISPR/Cas system beyond gene editing. *Biology Methods and Protocols,* 3.

ANTONY, P., ROSE, M., HEIDENREICH, A., KNUCHEL, R., GAISA, N. T. & DAHL, E. 2014. Epigenetic inactivation of ST6GAL1 in human bladder cancer. *BMC Cancer,* 14**,** 901.

ANUGRAHAM, M., JACOB, F., NIXDORF, S., EVEREST-DASS, A. V., HEINZELMANN-SCHWARZ, V. & PACKER, N. H. 2014. Specific glycosylation of membrane proteins in epithelial ovarian cancer cell lines: glycan structures reflect gene expression and DNA methylation status. *Mol Cell Proteomics,* 13**,** 2213-32.

APPENHEIMER, M. M., HUANG, R.-Y., CHANDRASEKARAN, E. V., DALZIEL, M., HU, Y. P., SOLOWAY, P. D., WUENSCH, S. A., MATTA, K. L. & LAU, J. T. Y. 2003a. Biological contribution of P1 promoter–mediated expression of ST6Gal I sialyltransferase. *Glycobiology,* 13**,** 591-600.

APPENHEIMER, M. M., HUANG, R. Y., CHANDRASEKARAN, E. V., DALZIEL, M., HU, Y. P., SOLOWAY, P. D., WUENSCH, S. A., MATTA, K. L. & LAU, J. T. 2003b. Biologic contribution of P1 promoter-mediated expression of ST6Gal I sialyltransferase. *Glycobiology,* 13**,** 591-600.

ARMSTRONG, R. T., KUSHNIR, A. S. & WHITE, J. M. 2000. The transmembrane domain of influenza hemagglutinin exhibits a stringent length requirement to support the hemifusion to fusion transition. *J Cell Biol,* 151**,** 425-37.

AWATA, H., WATANABE, T., HAMANAKA, Y., MITO, T., NOJI, S. & MIZUNAMI, M. 2015. Knockout crickets for the study of learning and memory: Dopamine receptor Dop1 mediates aversive but not appetitive reinforcement in crickets. *Sci Rep,* 5**,** 15885.

AZOULAY-DUPUIS, E., LAMBRE, C. R., SOLER, P., MOREAU, J. & THIBON, M. 1984. Lung alterations in guinea-pigs infected with influenza virus. *J Comp Pathol,* 94**,** 273-83.

BANNISTER, S. C., WISE, T. G., CAHILL, D. M. & DORAN, T. J. 2007. Comparison of chicken 7SK and U6 RNA polymerase III promoters for short hairpin RNA expression. *BMC Biotechnol.*

BEVAN, M., BARNES, W. M. & CHILTON, M. D. 1983. Structure and transcription of the nopaline synthase gene region of T-DNA. *Nucleic Acids Res,* 11**,** 369-85.

BI, Y., HUA, Z., LIU, X., HUA, W., REN, H., XIAO, H., ZHANG, L., LI, L., WANG, Z., LAIBLE, G., WANG, Y., DONG, F. & ZHENG, X. 2016. Isozygous and selectable marker-free MSTN knockout cloned pigs generated by the combined use of CRISPR/Cas9 and Cre/LoxP. *Scientific Reports,* 6**,** 31729.

BIBIKOVA, M., GOLIC, M., GOLIC, K. G. & CARROLL, D. 2002. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics,* 161**,** 1169-75.

BIRD, A. M. D. A. A. 2011. *CpG islands and the regulation of transcription Aimée M. Deaton and Adrian Bird1* [Online]. [Accessed].

BOHM, S., SCHWAB, I., LUX, A. & NIMMERJAHN, F. 2012. The role of sialic acid as a modulator of the anti-inflammatory activity of IgG. *Semin Immunopathol,* 34**,** 443-53.

BOLTZ, D. A., ALDRIDGE, J. R., WEBSTER, R. G. & GOVORKOVA, E. A. 2010. Drugs in Development for Influenza. *Drugs,* 70**,** 1349-62.

BOSSELMAN, R. A., HSU, R. Y., BOGGS, T., HU, S., BRUSZEWSKI, J., OU, S., KOZAR, L., MARTIN, F., GREEN, C., JACOBSEN, F. & ET AL. 1989. Germline transmission of exogenous genes in the chicken. *Science,* 243**,** 533-5.

BRIDGES, C. B., KUEHNERT, M. J. & HALL, C. B. 2003. Transmission of influenza: implications for control in health care settings. *Clin Infect Dis,* 37**,** 1094-101.

BRINKMAN, E. K., CHEN, T., AMENDOLA, M. & VAN STEENSEL, B. 2014. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res,* 42**,** e168.

BYRD-LEOTIS, L., CUMMINGS, R. D. & STEINHAUER, D. A. 2017. The Interplay between the Host Receptor and Influenza Virus Hemagglutinin and Neuraminidase. *Int J Mol Sci.*

CAMPBELL, K. H., MCWHIR, J., RITCHIE, W. A. & WILMUT, I. 1996. Sheep cloned by nuclear transfer from a cultured cell line. *Nature,* 380**,** 64-6.

CANNY, M. D., MOATTI, N., WAN, L. C. K., FRADET-TURCOTTE, A., KRASNER, D., MATEOS-GOMEZ, P. A., ZIMMERMANN, M., ORTHWEIN, A., JUANG, Y. C., ZHANG, W., NOORDERMEER, S. M., SECLEN, E., WILSON, M. D., VOROBYOV, A., MUNRO, M., ERNST, A., NG, T. F., CHO, T., CANNON, P. M., SIDHU, S. S., SICHERI, F. & DUROCHER, D. 2018. Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR-Cas9 genome-editing efficiency. *Nat Biotechnol,* 36**,** 95-102.

CARLSON, D. F., TAN, W., LILLICO, S. G., STVERAKOVA, D., PROUDFOOT, C., CHRISTIAN, M., VOYTAS, D. F., LONG, C. R., WHITELAW, C. B. & FAHRENKRUG, S. C. 2012. Efficient TALEN-mediated gene knockout in livestock. *Proc Natl Acad Sci U S A,* 109**,** 17382-7.

CHANDRASEGARAN, S. & SMITH, J. 1999. Chimeric restriction enzymes: what is next? *Biol Chem,* 380**,** 841-8.

CHARLES W. HUTTON, A. P. C., JOHN R. CLAMP,. 1987. *The gut in the acute phase response: Changes in colonic and hepatic enzyme activity in response to dermal inflammation in the rat* [Online]. [Accessed].

CHAVAN, R. D., SHINDE, P., GIRKAR, K., MADAGE, R. & CHOWDHARY, A. 2016. Assessment of Anti-Influenza Activity and Hemagglutination Inhibition of Plumbago indica and Allium sativum Extracts. *Pharmacognosy Res,* 8**,** 105-11.

CHEN, B., GILBERT, L. A., CIMINI, B. A., SCHNITZBAUER, J., ZHANG, W., LI, G. W., PARK, J., BLACKBURN, E. H., WEISSMAN, J. S., QI, L. S. & HUANG, B. 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell,* 155**,** 1479-91.

CHENG, E., VAISICA, J. A., OU, J., BARYSHNIKOVA, A., LU, Y., ROTH, F. P. & BROWN, G. W. 2012. Genome Rearrangements Caused by Depletion of Essential DNA Replication Proteins in Saccharomyces cerevisiae. *Genetics.*

CHIRUVELLA, K. K., LIANG, Z. & WILSON, T. E. 2013. Repair of Double-Strand Breaks by End Joining. *Cold Spring Harb Perspect Biol.*

CHO, S. W., KIM, S., KIM, J. M. & KIM, J. S. 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol,* 31**,** 230-2.

CHRISTINE BURKARD, S. G. L., ELIZABETH REID, BEN JACKSON, ALAN J. MILEHAM, TAHAR AIT-ALI, C. BRUCE A. WHITELAW, ALAN L. ARCHIBALD 2017. Precision engineering for PRRSV resistance in pigs: Macrophages from genome edited pigs lacking CD163 SRCR5 domain are fully resistant to both PRRSV genotypes while maintaining biological function.

CHU, V. C. & WHITTAKER, G. R. 2004. Influenza virus entry and infection require host cell N-linked glycoprotein.

CHU, V. T., WEBER, T., WEFERS, B., WURST, W., SANDER, S., RAJEWSKY, K. & KÜHN, R. 2015. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature Biotechnology,* 33**,** 543.

COLLINS, B. E., SMITH, B. A., BENGTSON, P. & PAULSON, J. C. 2006. Ablation of CD22 in ligand-deficient mice restores B cell receptor signaling. *Nat Immunol,* 7**,** 199-206.

COLOSIMO, A., GONCZ, K. K., HOLMES, A. R., KUNZELMANN, K., NOVELLI, G., MALONE, R. W., BENNETT, M. J. & GRUENERT, D. C. 2000. Transfer and expression of foreign genes in mammalian cells. *Biotechniques,* 29**,** 314-8, 320-2, 324 passim.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science,* 339**,** 819-23.

CORNU, T. I., THIBODEAU-BEGANNY, S., GUHL, E., ALWIN, S., EICHTINGER, M., JOUNG, J. K. & CATHOMEN, T. 2008. DNA-binding Specificity Is a Major Determinant of the Activity and Toxicity of Zinc-finger Nucleases. *Mol Ther,* 16**,** 352-358.

COSTANTINI, S., WOODBINE, L., ANDREOLI, L., JEGGO, P. A. & VINDIGNI, A. 2007. Interaction of the Ku heterodimer with the DNA ligase IV/Xrcc4 complex and its regulation by DNA-PK. *DNA Repair (Amst),* 6**,** 712-22.

DABNEY, J., MEYER, M. & PAABO, S. 2013. Ancient DNA damage. *Cold Spring Harb Perspect Biol,* 5.

DAI, Y., VAUGHT, T. D., BOONE, J., CHEN, S. H., PHELPS, C. J., BALL, S., MONAHAN, J. A., JOBST, P. M., MCCREATH, K. J., LAMBORN, A. E., COWELL-LUCERO, J. L., WELLS, K. D., COLMAN, A., POLEJAEVA, I. A. & AYARES, D. L. 2002. Targeted disruption of the alpha1,3-galactosyltransferase gene in cloned pigs. *Nat Biotechnol,* 20**,** 251-5.

DALL'OLIO, F. 2000. The sialyl-alpha2,6-lactosaminyl-structure: biosynthesis and functional role. *Glycoconj J,* 17**,** 669-76.

DALL'OLIO, F., CHIRICOLO, M., CECCARELLI, C., MINNI, F., MARRANO, D. & SANTINI, D. 2000. Beta-galactoside alpha2,6 sialyltransferase in human colon cancer: contribution of multiple transcripts to regulation of enzyme activity and reactivity with Sambucus nigra agglutinin. *Int J Cancer,* 88**,** 58-65.

DALL'OLIO, F., CHIRICOLO, M. & LAU, J. T. 1999. Differential expression of the hepatic transcript of beta-galactoside alpha2,6-sialyltransferase in human colon cancer cell lines. *Int J Cancer,* 81**,** 243-7.

DALZIEL, M., HUANG, R. Y., DALL'OLIO, F., MORRIS, J. R., TAYLOR-PAPADIMITRIOU, J. & LAU, J. T. Y. 2001. Mouse ST6Gal sialyltransferase gene expression during mammary gland lactation. *Glycobiology,* 11**,** 407-412.

DATTA, A. K. & PAULSON, J. C. 1997. Sialylmotifs of sialyltransferases. *Indian J Biochem Biophys,* 34**,** 157-65.

DAWSON, W. K., BIO-INFORMATION LAB, U. O. T., UNIVERSITY OF WARSAW, C. O. N. T. C., WARSAW, POLAND, LAZNIEWSKI, M., UNIVERSITY OF WARSAW, C. O. N. T. C., WARSAW, POLAND, PLEWCZYNSKI, D. & UNIVERSITY OF WARSAW, C. O. N. T. C., WARSAW, POLAND 2017. RNA structure interactions and ribonucleoprotein processes of the influenza A virus. *Briefings in Functional Genomics,* 17**,** 402-414.

DE WIT, E., SPRONKEN, M. I., BESTEBROER, T. M., RIMMELZWAAN, G. F., OSTERHAUS, A. D. & FOUCHIER, R. A. 2004. Efficient generation and growth of influenza virus A/PR/8/34 from eight cDNA fragments. *Virus Res,* 103**,** 155-61.

DELGADO-ORTEGA, M., MELO, S., PUNYADARSANIYA, D., RAMÉ, C., OLIVIER, M., SOUBIEUX, D., MARC, D., SIMON, G., HERRLER, G., BERRI, M., DUPONT, J. & MEURENS, F. 2014. Innate immune response to a H3N2 subtype swine influenza virus in newborn porcine trachea cells, alveolar macrophages, and precision-cut lung slices. *Vet Res.*

DELTCHEVA, E., CHYLINSKI, K., SHARMA, C. M., GONZALES, K., CHAO, Y., PIRZADA, Z. A., ECKERT, M. R., VOGEL, J. & CHARPENTIER, E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature,* 471**,** 602-7.

DERIANO, L. & ROTH, D. B. 2013. Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. *Annu Rev Genet,* 47**,** 433-55.

DIEZ-VILLASENOR, C., ALMENDROS, C., GARCIA-MARTINEZ, J. & MOJICA, F. J. 2010. Diversity of CRISPR loci in Escherichia coli. *Microbiology,* 156**,** 1351-1361.

DIGARD, P., DUTIA, B., TURNBULL, M. L., BIOTECHNOLOGY & BIOLOGICAL SCIENCES RESEARCH, C. 2017. *Role of the NS segment of Influenza A virus in setting host range and pathogenicity.* PhD Doctor of Philosophy, The University of Edinburgh.

DOMINGUEZ, A. A., LIM, W. A. & QI, L. S. 2016. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat Rev Mol Cell Biol,* 17**,** 5-15.

DORNER, L. F., BITINAITE, J., WHITAKER, R. D. & SCHILDKRAUT, I. 1999. Genetic analysis of the base-specific contacts of BamHI restriction endonuclease. *J Mol Biol,* 285**,** 1515-23.

DOWDLE, W. R. 2006. Influenza Pandemic Periodicity, Virus Recycling, and the Art of Risk Assessment. *Emerg Infect Dis.*

DOYON, Y., MCCAMMON, J. M., MILLER, J. C., FARAJI, F., NGO, C., KATIBAH, G. E., AMORA, R., HOCKING, T. D., ZHANG, L., REBAR, E. J., GREGORY, P. D., URNOV, F. D. & AMACHER, S. L. 2008. Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol,* 26**,** 702-8.

DUBCHAK, I. & FRAZER, K. 2003. Multi-species sequence comparison: the next frontier in genome annotation. *Genome Biol.*

E.M.SOUTHERN 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis Author links open overlay panel. *Journal of Molecular Biology*.

EDINGER, T. O., POHL, M. O. & STERTZ, S. 2014. Entry of influenza A virus: host factors and antiviral targets. *J Gen Virol,* 95**,** 263-77.

EID, A. & MAHFOUZ, M. M. 2016. Genome editing: the road of CRISPR/Cas9 from bench to clinic. *Exp Mol Med,* 48**,** e265.

ENGLUND, J. A. 2002. Antiviral therapy of influenza. *Semin Pediatr Infect Dis,* 13**,** 120-8.

FAURE, G., OGURTSOV, A. Y., SHABALINA, S. A. & KOONIN, E. V. 2016. Role of mRNA structure in the control of protein folding. *Nucleic Acids Res,* 44**,** 10898-911.

FERRARI, M., SCALVINI, A., LOSIO, M. N., CORRADI, A., SONCINI, M., BIGNOTTI, E., MILANESI, E., AJMONE-MARSAN, P., BARLATI, S., BELLOTTI, D. & TONELLI, M. 2003. Establishment and characterization of two new pig cell lines for use in virological diagnostic laboratories. *Journal of Virological Methods,* 107**,** 205-212.

FISCHER, E. & BROSSMER, R. 1995. Sialic acid-binding lectins: submolecular specificity and interaction with sialoglycoproteins and tumour cells. *Glycoconj J,* 12**,** 707-13.

FISCHER, K. M. 1992. Transgenic domestic animals provide an animal model for rheumatoid arthritis. *Med Hypotheses,* 38**,** 240-3.

GAJ, T., GERSBACH, C. A. & BARBAS, C. F. 2013. ZFN, TALEN and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol,* 31**,** 397-405.

GARNEAU, J. E., DUPUIS, M.-È., VILLION, M., ROMERO, D. A., BARRANGOU, R., BOYAVAL, P., FREMAUX, C., HORVATH, P., MAGADÁN, A. H. & MOINEAU, S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature,* 468**,** 67.

GARTEN, R. J., DAVIS, C. T., RUSSELL, C. A., SHU, B., LINDSTROM, S., BALISH, A., SESSIONS, W. M., XU, X., SKEPNER, E., DEYDE, V., OKOMO-ADHIAMBO, M., GUBAREVA, L., BARNES, J., SMITH, C. B., EMERY, S. L., HILLMAN, M. J., RIVAILLER, P., SMAGALA, J., DE GRAAF, M., BURKE, D. F., FOUCHIER, R. A., PAPPAS, C., ALPUCHE-ARANDA, C. M., LOPEZ-GATELL, H., OLIVERA, H., LOPEZ, I., MYERS, C. A., FAIX, D., BLAIR, P. J., YU, C., KEENE, K. M., DOTSON, P. D., JR., BOXRUD, D., SAMBOL, A. R., ABID, S. H., ST GEORGE, K., BANNERMAN, T.,

MOORE, A. L., STRINGER, D. J., BLEVINS, P., DEMMLER-HARRISON, G. J., GINSBERG, M., KRINER, P., WATERMAN, S., SMOLE, S., GUEVARA, H. F., BELONGIA, E. A., CLARK, P. A., BEATRICE, S. T., DONIS, R., KATZ, J., FINELLI, L., BRIDGES, C. B., SHAW, M., JERNIGAN, D. B., UYEKI, T. M., SMITH, D. J., KLIMOV, A. I. & COX, N. J. 2009. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science,* 325**,** 197-201.

GATES, K. S. 2009. An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals. *Chem Res Toxicol,* 22**,** 1747-60.

GEURTS, A. M., COST, G. J., FREYVERT, Y., ZEITLER, B., MILLER, J. C., CHOI, V. M., JENKINS, S. S., WOOD, A., CUI, X., MENG, X., VINCENT, A., LAM, S., MICHALKIEWICZ, M., SCHILLING, R., FOECKLER, J., KALLOWAY, S., WEILER, H., MÉNORET, S., ANEGON, I., DAVIS, G. D., ZHANG, L., REBAR, E. J., GREGORY, P. D., URNOV, F. D., JACOB, H. J. & BUELOW, R. 2009. Knockout Rats Produced Using Designed Zinc Finger Nucleases. *Science,* 325**,** 433.

GLASER, L., STEVENS, J., ZAMARIN, D., WILSON, I. A., GARCÍA-SASTRE, A., TUMPEY, T. M., BASLER, C. F., TAUBENBERGER, J. K. & PALESE, P. 2005. A Single Amino Acid Substitution in 1918 Influenza Virus Hemagglutinin Changes Receptor Binding Specificity. *J Virol.*

GORDON, J. W. & RUDDLE, F. H. 1981. Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science,* 214**,** 1244-6.

GORDON, J. W., SCANGOS, G. A., PLOTKIN, D. J., BARBOSA, J. A. & RUDDLE, F. H. 1980. Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc Natl Acad Sci U S A,* 77**,** 7380-4.

GRETSCHEL, S., KEMMNER, W., FISCHER, J. & SCHLAG, P. M. 1998. [Significance of alpha 2,6-specific sialyltransferase ST6N and alpha 2,3-specific sialyltransferase ST3N in stomach carcinoma]. *Langenbecks Arch Chir Suppl Kongressbd,* 115**,** 475-8.

GROHSKOPF, L. A., OLSEN, S. J., SOKOLOW, L. Z., BRESEE, J. S., COX, N. J., BRODER, K. R., KARRON, R. A. & WALTER, E. B. 2014. Prevention and control of seasonal influenza with vaccines: recommendations of the Advisory Committee on Immunization Practices (ACIP) -- United States, 2014-15 influenza season. *MMWR Morb Mortal Wkly Rep,* 63**,** 691-7.

GUO, Q., MINTIER, G., MA-EDMONDS, M., STORTON, D., WANG, X., XIAO, X., KIENZLE, B., ZHAO, D. & FEDER, J. N. 2018. 'Cold shock' increases the frequency of homology directed repair gene editing in induced pluripotent stem cells. *Scientific Reports,* 8**,** 2080.

GUTSCHNER, T., HAEMMERLE, M., GENOVESE, G., DRAETTA, G. F. & CHIN, L. 2016. Post-translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair. *Cell Rep,* 14**,** 1555-1566.

HAI, T., TENG, F., GUO, R., LI, W. & ZHOU, Q. 2014. One-step generation of knockout pigs by zygote injection of CRISPR/Cas system. *Cell Research,* 24**,** 372.

HAMAMOTO, T., KAWASAKI, M., KUROSAWA, N., NAKAOKA, T., LEE, Y. C. & TSUJI, S. 1993. Two step single primer mediated polymerase chain reaction. Application to cloning of putative mouse, beta-galactoside alpha 2,6-sialyltransferase cDNA. *Bioorg Med Chem,* 1**,** 141-5.

HAMILTON, B. S., WHITTAKER, G. R. & DANIEL, S. 2012. Influenza Virus-Mediated Membrane Fusion: Determinants of Hemagglutinin Fusogenic Activity and Experimental Approaches for Assessing Virus Fusion. *Viruses.*

HAMMER, R. E., PURSEL, V. G., JR, C. E. R., WALL, R. J., BOLT, D. J., EBERT, K. M., PALMITER, R. D. & BRINSTER, R. L. 1985a. Production of transgenic rabbits, sheep and pigs by microinjection. *Nature,* 315**,** 680.

HAMMER, R. E., PURSEL, V. G., REXROAD, C. E., JR., WALL, R. J., BOLT, D. J., EBERT, K. M., PALMITER, R. D. & BRINSTER, R. L. 1985b. Production of transgenic rabbits, sheep and pigs by microinjection. *Nature,* 315**,** 680-3.

HARADA, N., UTSUMI, T. & TAKAGI, Y. 1993. Tissue-specific expression of the human aromatase cytochrome P-450 gene by alternative use of multiple exons 1 and

promoters, and switching of tissue-specific exons 1 in carcinogenesis. *Proc Natl Acad Sci U S A,* 90**,** 11312-6.
HAUSCHILD, J., PETERSEN, B., SANTIAGO, Y., QUEISSER, A. L., CARNWATH, J. W., LUCAS-HAHN, A., ZHANG, L., MENG, X., GREGORY, P. D., SCHWINZER, R., COST, G. J. & NIEMANN, H. 2011. Efficient generation of a biallelic knockout in pigs using zinc-finger nucleases. *Proc Natl Acad Sci U S A.*
HEIDER, A., MOCHALOVA, L., HARDER, T., TUZIKOV, A., BOVIN, N., WOLFF, T., MATROSOVICH, M., SCHWEIGER, B. & GARCÍA-SASTRE, A. 2015. Alterations in Hemagglutinin Receptor-Binding Specificity Accompany the Emergence of Highly Pathogenic Avian Influenza Viruses.
HELEN M. WISE, E. C. H., BRETT W. JAGGER, AMANDA D. STUART, ZI H. KANG, NICOLE ROBB, LOUIS M. SCHWARTZMAN, JOHN C. KASH, ERVIN FODOR, ANDREW E. FIRTH, JULIA R. GOG, JEFFERY K. TAUBENBERGER, PAUL DIGARD 2012. Identification of a Novel Splice Variant Form of the Influenza A Virus M2 Ion Channel with an Antigenically Distinct Ectodomain.
HENNET, T., CHUI, D., PAULSON, J. C. & MARTH, J. D. 1998. Immune regulation by the ST6Gal sialyltransferase. *Proc Natl Acad Sci U S A,* 95**,** 4504-9.
HOCKEL, M., SCHLENGER, K., ARAL, B., MITZE, M., SCHAFFER, U. & VAUPEL, P. 1996. Association between tumor hypoxia and malignant progression in advanced cancer of the uterine cervix. *Cancer Res,* 56**,** 4509-15.
HOCKEMEYER, D., WANG, H., KIANI, S., LAI, C. S., GAO, Q., CASSADY, J. P., COST, G. J., ZHANG, L., SANTIAGO, Y., MILLER, J. C., ZEITLER, B., CHERONE, J. M., MENG, X., HINKLEY, S. J., REBAR, E. J., GREGORY, P. D., URNOV, F. D. & JAENISCH, R. 2011. Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol,* 29**,** 731-4.
HU, Y., LU, X., ZHOU, G., BARNES, E. L. & LUO, G. 2009. Recql5 Plays an Important Role in DNA Replication and Cell Survival after Camptothecin Treatment. *Mol Biol Cell.*
IBRICEVIC, A., PEKOSZ, A., WALTER, M. J., NEWBY, C., BATTAILE, J. T., BROWN, E. G., HOLTZMAN, M. J. & BRODY, S. L. 2006. Influenza Virus Receptor Specificity and Cell Tropism in Mouse and Human Airway Epithelial Cells.
IIJIMA, R., TAKAHASHI, H., IKEGAMI, S. & YAMAZAKI, M. 2007. Characterization of the Reaction between Sialic Acid (N-Acetylneuraminic Acid) and Hydrogen Peroxide. *Biol Pharm Bull,* 30**,** 580-2.
IMAMAKI, R., OGAWA, K., KIZUKA, Y., KOMI, Y., KOJIMA, S., KOTANI, N., HONKE, K., HONDA, T., TANIGUCHI, N. & KITAZUME, S. 2018. Glycosylation controls cooperative PECAM-VEGFR2-beta3 integrin functions at the endothelial surface for tumor angiogenesis. *Oncogene,* 37**,** 4287-4299.
IRONS, E. E. & LAU, J. T. Y. 2018. Systemic ST6Gal-1 Is a Pro-survival Factor for Murine Transitional B Cells. *Front Immunol,* 9.
IWATSUKI-HORIMOTO, K., NAKAJIMA, N., ICHIKO, Y., SAKAI-TAGAWA, Y., NODA, T., HASEGAWA, H. & KAWAOKA, Y. 2018. Syrian Hamster as an Animal Model for the Study of Human Influenza Virus Infection. *J Virol.*
JAENISCH, R., FAN, H. & CROKER, B. 1975. Infection of preimplantation mouse embryos and of newborn mice with leukemia virus: tissue distribution of viral DNA and RNA and leukemogenesis in the adult animal. *Proc Natl Acad Sci U S A,* 72**,** 4008-12.
JAENISCH, R. & MINTZ, B. 1974. Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA. *Proc Natl Acad Sci U S A,* 71**,** 1250-4.
JANG, Y. H. & SEONG, B. L. 2012. Principles underlying rational design of live attenuated influenza vaccines. *Clin Exp Vaccine Res,* 1**,** 35-49.
JANSEN, R., EMBDEN, J. D., GAASTRA, W. & SCHOULS, L. M. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol,* 43**,** 1565-75.
JIN, H. & CHEN, Z. 2014. Production of live attenuated influenza vaccines against seasonal and potential pandemic influenza viruses. *Curr Opin Virol,* 6**,** 34-9.
JIN, Y., XU, X., WANG, X., KUANG, H., OSTERMAN, M., FENG, S., HAN, D., WU, Y., LI, M. & GUO, H. 2016. Increasing sensitivity to DNA damage is a potential driver for human ovarian cancer. *Oncotarget,* 7**,** 49710-49721.

JONES, M. B., NASIRIKENARI, M., FENG, L., MIGLIORE, M. T., CHOI, K. S., KAZIM, L. & LAU, J. T. 2010. Role for hepatic and circulatory ST6Gal-1 sialyltransferase in regulating myelopoiesis. *J Biol Chem,* 285**,** 25009-17.

JONES, M. B., OSWALD, D. M., JOSHI, S., WHITEHEART, S. W., ORLANDO, R. & COBB, B. A. 2016. B-cell-independent sialylation of IgG. *Proc Natl Acad Sci U S A,* 113**,** 7207-12.

JONGKON, N., MOKMAK, W., CHUAKHEAW, D., SHAW, P. J., TONGSIMA, S. & SANGMA, C. 2009. Prediction of avian influenza A binding preference to human receptor using conformational analysis of receptor bound to hemagglutinin. *BMC Genomics,* 10.

JULIANNA HAN, J. T. P., CINDY CHEN, JORGE ANDRADE, & BENJAMIN TENOEVER, B. M. 2018. Genome-wide CRISPR/Cas9 Screen Identifies Host Factors Essential for Influenza Virus Replication: Cell Reports.

KAPLAN HA , E. A. 1983. Studies on the effect of inflammation on rat liver and serum sialyltransferase. Evidence that inflammation causes release of Gal beta 1 leads to 4G... - PubMed - NCBI.

KIM, D., BAE, S., PARK, J., KIM, E., KIM, S., YU, H. R., HWANG, J., KIM, J.-I. & KIM, J.-S. 2015. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nature Methods,* 12**,** 237.

KITAZUME, S. 2014. ST6 Beta-Galactoside Alpha-2,6-Sialyltranferase 1 (ST6GAL1) | SpringerLink.

KLIMOV, A., BALISH, A., VEGUILLA, V., SUN, H., SCHIFFER, J., LU, X., KATZ, J. M. & HANCOCK, K. 2012. Influenza virus titration, antigenic characterization, and serological methods for antibody detection. *Methods Mol Biol,* 865**,** 25-51.

KOSICKI, M., TOMBERG, K. & BRADLEY, A. 2018. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol,* 36**,** 765-771.

KOZAK, M. 1984. Selection of initiation sites by eucaryotic ribosomes: effect of inserting AUG triplets upstream from the coding sequence for preproinsulin. *Nucleic Acids Res,* 12**,** 3873-93.

KREIJTZ, J. H., FOUCHIER, R. A. & RIMMELZWAAN, G. F. 2011. Immune responses to influenza virus infection. *Virus Res,* 162**,** 19-30.

KRIMPENFORT, P., RADEMAKERS, A., EYESTONE, W., VAN DER SCHANS, A., VAN DEN BROEK, S., KOOIMAN, P., KOOTWIJK, E., PLATENBURG, G., PIEPER, F., STRIJKER, R. & ET AL. 1991. Generation of transgenic dairy cattle using 'in vitro' embryo production. *Biotechnology (N Y),* 9**,** 844-7.

KUCHIPUDI, S. V., NELLI, R., WHITE, G. A., BAIN, M., CHANG, K. C. & DUNHAM, S. 2009. Differences in influenza virus receptors in chickens and ducks: Implications for interspecies transmission. *J Mol Genet Med,* 3**,** 143-51.

KUMLIN, U., OLOFSSON, S., DIMOCK, K. & ARNBERG, N. 2008. Sialic acid tissue distribution and influenza virus tropism. *Influenza Other Respir Viruses.*

KUROSAWA, N., KAWASAKI, M., HAMAMOTO, T., NAKAOKA, T., LEE, Y. C., ARITA, M. & TSUJI, S. 1994. Molecular cloning and expression of chick embryo Gal beta 1,4GlcNAc alpha 2,6-sialyltransferase. Comparison with the mammalian enzyme. *Eur J Biochem,* 219**,** 375-81.

LANGERAK, P., MEJIA-RAMIREZ, E., LIMBO, O. & RUSSELL, P. 2011. Release of Ku and MRN from DNA ends by Mre11 nuclease activity and Ctp1 is required for homologous recombination repair of double-strand breaks. *PLoS Genet,* 7**,** e1002271.

LANGERAK, P. & RUSSELL, P. 2011. Regulatory networks integrating cell cycle control with DNA damage checkpoints and double-strand break repair. *Philos Trans R Soc Lond B Biol Sci,* 366**,** 3562-71.

LAZNIEWSKI, M., DAWSON, W. K., SZCZEPINSKA, T. & PLEWCZYNSKI, D. 2018. The structural variability of the influenza A hemagglutinin receptor-binding site. *Brief Funct Genomics,* 17**,** 415-427.

LI, N., QI, Y., ZHANG, F. Y., YU, X. H., WU, Y. G., CHEN, Y., JIANG, C. L. & KONG, W. 2011. Overexpression of alpha-2,6 sialyltransferase stimulates propagation of human influenza viruses in Vero cells. *Acta Virol,* 55**,** 147-53.

LI, X., YANG, Y., BU, L., GUO, X., TANG, C., SONG, J., FAN, N., ZHAO, B., OUYANG, Z., LIU, Z., ZHAO, Y., YI, X., QUAN, L., LIU, S., YANG, Z., OUYANG, H., CHEN, Y. E.,

WANG, Z. & LAI, L. 2014. Rosa26-targeted swine models for stable gene over-expression and Cre-mediated lineage tracing. *Cell Res.* England.
LIAO, S., TAMMARO, M. & YAN, H. 2016. The structure of ends determines the pathway choice and Mre11 nuclease dependency of DNA double-strand break repair. *Nucleic Acids Res,* 44**,** 5689-701.
LIEBER, M. R. 2010. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem,* 79**,** 181-211.
LILLICO, S. G., PROUDFOOT, C., CARLSON, D. F., STVERAKOVA, D., NEIL, C., BLAIN, C., KING, T. J., RITCHIE, W. A., TAN, W., MILEHAM, A. J., MCLAREN, D. G., FAHRENKRUG, S. C. & WHITELAW, C. B. 2013. Live pigs produced from genome edited zygotes. *Sci Rep,* 3**,** 2847.
LILLICO, S. G., PROUDFOOT, C., KING, T. J., TAN, W., ZHANG, L., MARDJUKI, R., PASCHON, D. E., REBAR, E. J., URNOV, F. D., MILEHAM, A. J., MCLAREN, D. G. & WHITELAW, C. B. A. 2016. Mammalian interspecies substitution of immune modulatory alleles by genome editing. *Scientific Reports,* 6**,** 21645.
LIN, S., STAAHL, B. T., ALLA, R. K. & DOUDNA, J. A. 2014. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife.*
LIN, S. C., KAPPES, M. A., CHEN, M. C., LIN, C. C. & WANG, T. T. 2017. Distinct susceptibility and applicability of MDCK derivatives for influenza virus research. *PLoS One.*
LIU, J., EHMSEN, K. T., HEYER, W. D. & MORRICAL, S. W. 2011. Presynaptic Filament Dynamics in Homologous Recombination and DNA Repair. *Crit Rev Biochem Mol Biol,* 46**,** 240-70.
LIU, S., WANG, Q., YU, X., LI, Y., GUO, Y., LIU, Z., SUN, F., HOU, W., LI, C., WU, L., GUO, D. & CHEN, S. 2018. HIV-1 inhibition in cells with CXCR4 mutant genome created by CRISPR-Cas9 and piggyBac recombinant technologies. *Scientific Reports,* 8**,** 8573.
LO, N. W. & LAU, J. T. 1996. Transcription of the beta-galactoside alpha 2,6-sialyltransferase gene in B lymphocytes is directed by a separate and distinct promoter. *Glycobiology,* 6**,** 271-9.
LO, N. W. & LAU, J. T. 1999. Transcription of the beta-galactoside alpha2,6-sialyltransferase gene (SIAT1) in B-lymphocytes: cell type-specific expression correlates with presence of the divergent 5'-untranslated sequence. *Glycobiology,* 9**,** 907-14.
MA, W., KAHN, R. E. & RICHT, J. A. 2009. The pig as a mixing vessel for influenza viruses: Human and veterinary implications. *J Mol Genet Med,* 3**,** 158-66.
MAIR, C. M., LUDWIG, K., HERRMANN, A. & SIEBEN, C. 2014. Receptor binding and pH stability - how influenza A virus hemagglutinin affects host-specific virus infection. *Biochim Biophys Acta,* 1838**,** 1153-68.
MAKSIMOVIC, J., CENTRE FOR REPRODUCTION AND DEVELOPMENT, M. I. O. M. R., CLAYTON 3168, AUSTRALIA, BIOSCIENCES RESEARCH DIVISION, D. O. P. I., BUNDOORA 3083, AUSTRALIA, SHARP, J. A., INSTITUTE FOR TECHNOLOGY RESEARCH AND INNOVATION, D. U., GEELONG 3214, AUSTRALIA, NICHOLAS, K. R., INSTITUTE FOR TECHNOLOGY RESEARCH AND INNOVATION, D. U., GEELONG 3214, AUSTRALIA, COCKS, B. G., BIOSCIENCES RESEARCH DIVISION, D. O. P. I., BUNDOORA 3083, AUSTRALIA, SAVIN, K. & BIOSCIENCES RESEARCH DIVISION, D. O. P. I., BUNDOORA 3083, AUSTRALIA 2010. Conservation of the ST6Gal I gene and its expression in the mammary gland. *Glycobiology,* 21**,** 467-481.
MAKSIMOVIC, J., CENTRE FOR REPRODUCTION AND DEVELOPMENT, M. I. O. M. R., CLAYTON 3168, AUSTRALIA, BIOSCIENCES RESEARCH DIVISION, D. O. P. I., BUNDOORA 3083, AUSTRALIA, SHARP, J. A., INSTITUTE FOR TECHNOLOGY RESEARCH AND INNOVATION, D. U., GEELONG 3214, AUSTRALIA, NICHOLAS, K. R., INSTITUTE FOR TECHNOLOGY RESEARCH AND INNOVATION, D. U., GEELONG 3214, AUSTRALIA, COCKS, B. G., BIOSCIENCES RESEARCH DIVISION, D. O. P. I., BUNDOORA 3083, AUSTRALIA, SAVIN, K. & BIOSCIENCES RESEARCH DIVISION, D. O. P. I., BUNDOORA 3083, AUSTRALIA 2011. Conservation of the ST6Gal I gene and its expression in the mammary gland. *Glycobiology,* 21**,** 467-481.

MANHARDT, C. T., PUNCH, P. R., DOUGHER, C. W. L. & LAU, J. T. Y. 2017. Extrinsic sialylation is dynamically regulated by systemic triggers in vivo. *J Biol Chem,* 292**,** 13514-13520.

MASCARENHAS, J. X., KOROKHOV, N., BURGER, L., KASSIM, A., TUTER, J., MILLER, D., BORGSCHULTE, T., GEORGE, H. J., CHANG, A., PINTEL, D. J., ONIONS, D. & KAYSER, K. J. 2017. Genetic engineering of CHO cells for viral resistance to minute virus of mice. *Biotechnol Bioeng,* 114**,** 576-588.

MATSUZAKI, Y., DEPARTMENT OF INFECTIOUS DISEASES, Y. U. S. O. M., IIDA-NISHI,, KATSUSHIMA, N., KATSUSHIMA PEDIATRIC CLINIC, M., NAGAI, Y., NAGAI CHILDREN'S CLINIC, M., SHOJI, M., SHOJI CLINIC, H.-M., ITAGAKI, T., YAMANOBE PEDIATRIC CLINIC, Y., SAKAMOTO, M., DEPARTMENT OF PEDIATRICS, Y. C. H. S., NANOKAMACHI, AND, KITAOKA, S., DEPARTMENT OF PEDIATRICS, T. K. H., MIYAGINO, AND, MIZUTA, K., YAMAGATA PREFECTURAL INSTITUTE OF PUBLIC HEALTH, T.-M., YAMAGATA, AND, NISHIMURA, H. & VIRUS RESEARCH CENTER, C. R. D., SENDAI MEDICAL CENTER, SENDAI, JAPAN 2006. Clinical Features of Influenza C Virus Infection in Children. *The Journal of Infectious Diseases,* 193**,** 1229-1235.

MCVEY, M. & LEE, S. E. 2008. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet,* 24**,** 529-38.

MENG, L., FOROUHAR, F., THIEKER, D., GAO, Z., RAMIAH, A., MONIZ, H., XIANG, Y., SEETHARAMAN, J., MILANINIA, S., SU, M., BRIDGER, R., VEILLON, L., AZADI, P., KORNHABER, G., WELLS, L., MONTELIONE, G. T., WOODS, R. J., TONG, L. & MOREMEN, K. W. 2013. Enzymatic basis for N-glycan sialylation: structure of rat alpha2,6-sialyltransferase (ST6GAL1) reveals conserved and unique features for glycan sialylation. *J Biol Chem,* 288**,** 34680-98.

MERCIER, D., WIERINCKX, A., OULMOUDEN, A., GALLET, P. F., PALCIC, M. M., HARDUIN-LEPERS, A., DELANNOY, P., PETIT, J. M., LEVEZIEL, H. & JULIEN, R. 1999. Molecular cloning, expression and exon/intron organization of the bovine beta-galactoside alpha2,6-sialyltransferase gene. *Glycobiology,* 9**,** 851-63.

MEURENS, F., SUMMERFIELD, A., NAUWYNCK, H., SAIF, L. & GERDTS, V. 2012. The pig: a model for human infectious diseases. *Trends Microbiol,* 20**,** 50-7.

MICHAELIS, M., DOERR, H. W. & CINATL, J., JR. 2009. An influenza A H1N1 virus revival - pandemic H1N1/09 virus. *Infection,* 37**,** 381-9.

MILLER, J., MCLACHLAN, A. D. & KLUG, A. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J,* 4**,** 1609-14.

MINAMI, A., SHIMONO, Y., MIZUTANI, K., NOBUTANI, K., MOMOSE, K., AZUMA, T. & TAKAI, Y. 2013. Reduction of the ST6 beta-galactosamide alpha-2,6-sialyltransferase 1 (ST6GAL1)-catalyzed sialylation of nectin-like molecule 2/cell adhesion molecule 1 and enhancement of ErbB2/ErbB3 signaling by microRNA-199a. *J Biol Chem,* 288**,** 11845-53.

MULLER, M., BRENIG, B., WINNACKER, E. L. & BREM, G. 1992. Transgenic pigs carrying cDNA copies encoding the murine Mx1 protein which confers resistance to influenza virus infection. *Gene,* 121**,** 263-70.

MUNSTER, V. J., DE WIT, E., VAN DEN BRAND, J. M., HERFST, S., SCHRAUWEN, E. J., BESTEBROER, T. M., VAN DE VIJVER, D., BOUCHER, C. A., KOOPMANS, M., RIMMELZWAAN, G. F., KUIKEN, T., OSTERHAUS, A. D. & FOUCHIER, R. A. 2009. Pathogenesis and transmission of swine-origin 2009 A(H1N1) influenza virus in ferrets. *Science,* 325**,** 481-3.

MWANGI, D. W. & BANSAL, D. D. 2003. N-glycolylneuraminic acid conjugates: implications of their absence in mammalian biochemistry. *Indian J Biochem Biophys,* 40**,** 217-25.

MÖRÖY, T. & HEYD, F. 2007. The impact of alternative splicing in vivo: Mouse models show the way.

NASIRIKENARI, M., CHANDRASEKARAN, E. V., MATTA, K. L., SEGAL, B. H., BOGNER, P. N., LUGADE, A. A., THANAVALA, Y., LEE, J. J. & LAU, J. T. Y. 2010. Altered eosinophil profile in mice with ST6Gal-1 deficiency: an additional role for ST6Gal-1 generated by the P1 promoter in regulating allergic inflammation. *J Leukoc Biol.*

NAYAK, D. P., BALOGUN, R. A., YAMADA, H., ZHOU, Z. H. & BARMAN, S. 2009. Influenza virus morphogenesis and budding. *Virus Res,* 143**,** 147-61.

NIU, D., WEI, H. J., LIN, L., GEORGE, H., WANG, T., LEE, I. H., ZHAO, H. Y., WANG, Y., KAN, Y., SHROCK, E., LESHA, E., WANG, G., LUO, Y., QING, Y., JIAO, D., ZHAO, H., ZHOU, X., WANG, S., WEI, H., GUELL, M., CHURCH, G. M. & YANG, L. 2017. Inactivation of porcine endogenous retrovirus in pigs using CRISPR-Cas9. *Science,* 357**,** 1303-1307.

O'NEILL, P. A., SHAABAN, A. M., WEST, C. R., DODSON, A., JARVIS, C., MOORE, P., DAVIES, M. P. A., SIBSON, D. R. & FOSTER, C. S. 2004. Increased risk of malignant progression in benign proliferating breast lesions defined by expression of heat shock protein 27. *Br J Cancer.*

OGATA, M., NAKAJIMA, M., KATO, T., OBARA, T., YAGI, H., KATO, K., USUI, T. & PARK, E. Y. 2009. Synthesis of sialoglycopolypeptide for potentially blocking influenza virus infection using a rat alpha2,6-sialyltransferase expressed in BmNPV bacmid-injected silkworm larvae. *BMC Biotechnol,* 9**,** 54.

ORTIGOZA, M. B., BLASER, S. B., ZAFAR, M. A., HAMMOND, A. J., WEISER, J. N., KLUGMAN, K. P., KRAMMER, F. & MCCULLERS, J. 2018. An Infant Mouse Model of Influenza Virus Transmission Demonstrates the Role of Virus-Specific Shedding, Humoral Immunity, and Sialidase Expression by Colonizing Streptococcus pneumoniae.

ORTIZ-SOTO, M. E. & SEIBEL, J. 2016. Expression of Functional Human Sialyltransferases ST3Gal1 and ST6Gal1 in Escherichia coli. *PLoS One,* 11**,** e0155410.

PAQUET, D., KWART, D., CHEN, A., SPROUL, A., JACOB, S., TEO, S., OLSEN, K. M., GREGG, A., NOGGLE, S. & TESSIER-LAVIGNE, M. 2016. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature,* 533**,** 125-9.

PARRISH, C. R., MURCIA, P. R. & HOLMES, E. C. 2015. Influenza virus reservoirs and intermediate hosts: dogs, horses, and new possibilities for influenza virus exposure of humans. *J Virol,* 89**,** 2990-4.

PATTANAYAK, V., LIN, S., GUILINGER, J. P., MA, E., DOUDNA, J. A. & LIU, D. R. 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol,* 31**,** 839-43.

PAVLETICH, N. P. & PABO, C. O. 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science,* 252**,** 809-17.

PEREZ, E. E., WANG, J., MILLER, J. C., JOUVENOT, Y., KIM, K. A., LIU, O., WANG, N., LEE, G., BARTSEVICH, V. V., LEE, Y. L., GUSCHIN, D. Y., RUPNIEWSKI, I., WAITE, A. J., CARPENITO, C., CARROLL, R. G., ORANGE, J. S., URNOV, F. D., REBAR, E. J., ANDO, D., GREGORY, P. D., RILEY, J. L., HOLMES, M. C. & JUNE, C. H. 2008. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol,* 26**,** 808-16.

PEREZ-PINERA, P., OUSTEROUT, D. G., BROWN, M. T. & GERSBACH, C. A. 2012. Gene targeting to the ROSA26 locus directed by engineered zinc finger nucleases. *Nucleic Acids Res,* 40**,** 3741-52.

PETIT, D., MIR, A. M., PETIT, J. M., THISSE, C., DELANNOY, P., ORIOL, R., THISSE, B. & HARDUIN-LEPERS, A. 2010. Molecular Phylogeny and Functional Genomics of β-Galactoside α2,6-Sialyltransferases That Explain Ubiquitous Expression of st6gal1 Gene in Amniotes*. *J Biol Chem.*

PETTERS, R. M., ALEXANDER, C. A., WELLS, K. D., COLLINS, E. B., SOMMER, J. R., BLANTON, M. R., ROJAS, G., HAO, Y., FLOWERS, W. L., BANIN, E., CIDECIYAN, A. V., JACOBSON, S. G. & WONG, F. 1997. Genetically engineered large animal model for studying cone photoreceptor survival and degeneration in retinitis pigmentosa. *Nat Biotechnol,* 15**,** 965-70.

PIETER MEYSMAN, J. C.-V., ENRIQUE MORETT, ROBERTO VIOLA, KRISTOF ENGELEN , KRIS LAUKENS 2014. Structural Properties of Prokaryotic Promoter Regions Correlate with Functional Features.

PLESCHKA, S. 2012. Overview of Influenza Viruses | SpringerLink. SpringerLink.

PORTELA, A. & DIGARD, P. 2002. The influenza virus nucleoprotein: a multifunctional RNA-binding protein pivotal to virus replication. *J Gen Virol,* 83**,** 723-34.

QIAN, Q., CHE, J., YE, L. & ZHONG, B. 2014. [The improvement and application of piggyBac transposon system in mammals]. *Yi Chuan,* 36**,** 965-73.

RACH, E. A., YUAN, H.-Y., MAJOROS, W. H., TOMANCAK, P. & OHLER, U. 2009. Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the Drosophila genome. *Genome Biology,* 10.

RAJAO, D. S., DANIELA S. RAJAO, D., PHD, IS A VISITING RESEARCH MICROBIOLOGIST POST-DOCTORAL ASSOCIATE, AND AMY L. VINCENT, DVM, PHD, IS A RESEARCH VETERINARY MEDICAL OFFICER IN THE VIRUS AND PRION RESEARCH UNIT AT THE NATIONAL ANIMAL DISEASE CENTER OF THE AGRICULTURAL RESEARCH SERVICE (ARS) OF THE UNITED STATES DEPARTMENT OF AGRICULTURE (USDA) IN AMES, IOWA., VINCENT, A. L. & DANIELA S. RAJAO, D., PHD, IS A VISITING RESEARCH MICROBIOLOGIST POST-DOCTORAL ASSOCIATE, AND AMY L. VINCENT, DVM, PHD, IS A RESEARCH VETERINARY MEDICAL OFFICER IN THE VIRUS AND PRION RESEARCH UNIT AT THE NATIONAL ANIMAL DISEASE CENTER OF THE AGRICULTURAL RESEARCH SERVICE (ARS) OF THE UNITED STATES DEPARTMENT OF AGRICULTURE (USDA) IN AMES, IOWA. 2011. Swine as a Model for Influenza A Virus Infection and Immunity. *ILAR Journal,* 56**,** 44-52.

RAN, Z., SHEN, H., LANG, Y., KOLB, E. A., TURAN, N., ZHU, L., MA, J., BAWA, B., LIU, Q., LIU, H., QUAST, M., SEXTON, G., KRAMMER, F., HAUSE, B. M., CHRISTOPHER-HENNINGS, J., NELSON, E. A., RICHT, J., LI, F., MA, W. & SANDRI-GOLDIN, R. M. 2015. Domestic Pigs Are Susceptible to Infection with Influenza B Viruses.

RICHARDSON, C. D., RAY, G. J., DEWITT, M. A., CURIE, G. L. & CORN, J. E. 2016. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol,* 34**,** 339-44.

RODGERS, K. & MCVEY, M. 2016. Error-Prone Repair of DNA Double-Strand Breaks. *J Cell Physiol,* 231**,** 15-24.

RODRIGUES, E. & MACAULEY, M. S. 2018. Hypersialylation in Cancer: Modulation of Inflammation and Therapeutic Opportunities. *Cancers (Basel).*

ROGERS, C. S., HAO, Y., ROKHLINA, T., SAMUEL, M., STOLTZ, D. A., LI, Y., PETROFF, E., VERMEER, D. W., KABEL, A. C., YAN, Z., SPATE, L., WAX, D., MURPHY, C. N., RIEKE, A., WHITWORTH, K., LINVILLE, M. L., KORTE, S. W., ENGELHARDT, J. F., WELSH, M. J. & PRATHER, R. S. 2008. Production of CFTR-null and CFTR-DeltaF508 heterozygous pigs by adeno-associated virus-mediated gene targeting and somatic cell nuclear transfer. *J Clin Invest,* 118**,** 1571-7.

RUBIN, G. M. & SPRADLING, A. C. 1982. Genetic transformation of Drosophila with transposable element vectors. *Science,* 218**,** 348-53.

RUIZ-HERNANDEZ, R., MWANGI, W., PEROVAL, M., SADEYEN, J. R., ASCOUGH, S., BALKISSOON, D., STAINES, K., BOYD, A., MCCAULEY, J., SMITH, A. & BUTTER, C. 2016. Host genetics determine susceptibility to avian influenza infection and transmission dynamics. *Sci Rep.*

SALLMYR, A. & TOMKINSON, A. E. 2018. Repair of DNA double-strand breaks by mammalian alternative end-joining pathways. *J Biol Chem,* 293**,** 10536-10546.

SALOMON, R. & WEBSTER, R. G. 2009. The Influenza Virus Enigma. *Cell,* 136**,** 402-10.

SAMJI, T. 2009. Influenza A: Understanding the Viral Life Cycle. *Yale J Biol Med,* 82**,** 153-9.

SAMY, A. & NAGUIB, M. M. 2018. Avian Respiratory Coinfection and Impact on Avian Influenza Pathogenicity in Domestic Poultry: Field and Experimental Findings. *Vet Sci.*

SANJANA, N. E., SHALEM, O. & ZHANG, F. 2014. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods,* 11**,** 783-4.

SCHNIEKE, A. E., KIND, A. J., RITCHIE, W. A., MYCOCK, K., SCOTT, A. R., RITCHIE, M., WILMUT, I., COLMAN, A. & CAMPBELL, K. H. 1997. Human factor IX transgenic sheep produced by transfer of nuclei from transfected fetal fibroblasts. *Science,* 278**,** 2130-3.

SCHROEDER, A., MUELLER, O., STOCKER, S., SALOWSKY, R., LEIBER, M., GASSMANN, M., LIGHTFOOT, S., MENZEL, W., GRANZOW, M. & RAGG, T. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol.*

SCHULTZ, M. J., HOLDBROOKS, A. T., CHAKRABORTY, A., GRIZZLE, W. E., LANDEN, C. N., BUCHSBAUM, D. J., CONNER, M. G., AREND, R. C., YOON, K. J., KLUG, C. A., BULLARD, D. C., KESTERSON, R. A., OLIVER, P. G., O'CONNOR, A. K., YODER,

B. K. & BELLIS, S. L. 2016a. The Tumor-Associated Glycosyltransferase ST6Gal-I Regulates Stem Cell Transcription Factors and Confers a Cancer Stem Cell Phenotype. *Cancer Res,* 76**,** 3978-88.
SCHULTZ, M. J., HOLDBROOKS, A. T., CHAKRABORTY, A., GRIZZLE, W. E., LANDEN, C. N., BUCHSBAUM, D. J., CONNER, M. G., AREND, R. C., YOON, K. J., KLUG, C. A., BULLARD, D. C., KESTERSON, R. A., OLIVER, P. G., O'CONNOR, A. K., YODER, B. K. & BELLIS, S. L. 2016b. The tumor-associated glycosyltransferase ST6Gal-I regulates stem cell transcription factors and confers a cancer stem cell phenotype. *Cancer Res,* 76**,** 3978-88.
SCOTT, S. P. & PANDITA, T. K. 2006. The Cellular Control of DNA Double-Strand Breaks. *J Cell Biochem,* 99**,** 1463-75.
SEKELSKY, J. 2017. DNA Repair in Drosophila: Mutagens, Models, and Missing Genes. *Genetics,* 205**,** 471-490.
SHAN, Q., WANG, Y., LI, J., ZHANG, Y., CHEN, K., LIANG, Z., ZHANG, K., LIU, J., XI, J. J., QIU, J. L. & GAO, C. 2013. Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol.* United States.
SHAO, W., LI, X., GORAYA, M. U., WANG, S. & CHEN, J. L. 2017. Evolution of Influenza A Virus by Mutation and Re-Assortment. *Int J Mol Sci,* 18.
SHINYA, K., EBINA, M., YAMADA, S., ONO, M., KASAI, N. & KAWAOKA, Y. 2006. Avian flu: influenza virus receptors in the human airway. *Nature,* 440**,** 435-6.
SHOPE, R. E. 1931. *SWINE INFLUENZA* [Online]. [Accessed].
SHYR, HSIEH, C.-C. & YI, M. 2011. Elevation of β-galactoside α2,6-sialyltransferase 1 in a fructose-responsive manner promotes pancreatic cancer metastasis.
SKEHEL, J. J. & WILEY, D. C. 2003. Receptor Binding and Membrane Fusion in Virus Entry: The Influenza Hemagglutinin. http://dx.doi.org/10.1146/annurev.biochem.69.1.531.
SLEMAN, S. S. 2018. How influenza a causes ''epidemics and pandemics'' among the population: novel targets for anti-influenza molecules.
SMITH, W., ANDREWES, C. H. & LAIDLAW, P. P. 1933. A VIRUS OBTAINED FROM INFLUENZA PATIENTS. *The Lancet,* 222**,** 66-68.
SPICUGLIA, S. & VANHILLE, L. 2012. Chromatin signatures of active enhancers. *Nucleus.*
SRIVENUGOPAL, K. S. & ALI-OSMAN, F. 2002. The DNA repair protein, O(6)-methylguanine-DNA methyltransferase is a proteolytic target for the E6 human papillomavirus oncoprotein. *Oncogene,* 21**,** 5940-5.
SRIWILAIJAROEN, N., KONDO, S., YAGI, H., TAKEMAE, N., SAITO, T., HIRAMATSU, H., KATO, K. & SUZUKI, Y. 2011. N-glycans from porcine trachea and lung: predominant NeuAcalpha2-6Gal could be a selective pressure for influenza variants in favor of human-type receptor. *PLoS One,* 6**,** e16302.
STANLEY N. COHEN, A. C. Y. C., AND LESLIE HSU 1972. Nonchromosomal Antibiotic Resistance in Bacteria: Genetic Transformation of Escherichia coli by R-Factor DNA Stanley N. Cohen, Annie C. Y. Chang, and Leslie Hsu.
SU, S., FU, X., LI, G., KERLIN, F. & VEIT, M. 2017. Novel Influenza D virus: Epidemiology, pathology, evolution and biological characteristics. *Virulence,* 8**,** 1580-91.
SUN, Y., BI, Y., PU, J., HU, Y., WANG, J., GAO, H., LIU, L., XU, Q., TAN, Y., LIU, M., GUO, X., YANG, H. & LIU, J. 2010. Guinea pig model for evaluating the potential public health risk of swine and avian influenza viruses. *PLoS One,* 5**,** e15537.
SUZUKI, O. & ABE, M. 2014. Galectin-1-mediated cell adhesion, invasion and cell death in human anaplastic large cell lymphoma: regulatory roles of cell surface glycans. *Int J Oncol,* 44**,** 1433-42.
SVENSSON, E. C., SOREGHAN, B., AND PAULSON, J.C. 1990. Organization of the beta-galactoside alpha 2, 6-sialyltransferase gene. Evidence for the transcriptional regulation of terminal glycosylation.
SWINDALL, A. F. & BELLIS, S. L. 2011. Sialylation of the Fas death receptor by ST6Gal-I provides protection against Fas-mediated apoptosis in colon carcinoma cells. *J Biol Chem,* 286**,** 22982-90.
SYMINGTON, L. S. 2014. End resection at double-strand breaks: mechanism and regulation. *Cold Spring Harb Perspect Biol,* 6.

T P O'HANLON, K. M. L., X C WANG AND J T LAU 1989. Tissue-specific expression of beta-galactoside alpha-2,6-sialyltransferase. Transcript heterogeneity predicts a divergent polypeptide.

TAKAHASHI, H., KATO, S., MURATA, M. & CARNINCI, P. 2012. CAGE- Cap Analysis Gene Expression: a protocol for the detection of promoter and transcriptional networks. *Methods Mol Biol,* 786**,** 181-200.

TAMERIUS, J., NELSON, M. I., ZHOU, S. Z., VIBOUD, C., MILLER, M. A. & ALONSO, W. J. 2011. Global Influenza Seasonality: Reconciling Patterns across Temperate and Tropical Regions. *Environ Health Perspect.*

TAUBENBERGER, J. K. & MORENS, D. M. 2013. Influenza Viruses: Breaking All the Rules. *mBio.*

TAYLOR, M. R. G., SPIREK, M., CHAURASIYA, K. R., WARD, J. D., CARZANIGA, R., YU, X., EGELMAN, E. H., COLLINSON, L. M., RUEDA, D., KREJCI, L. & BOULTON, S. J. 2015. Rad51 Paralogs Remodel Pre-synaptic Rad51 Filaments to Stimulate Homologous Recombination. *Cell,* 162**,** 271-286.

TURNER, M. W. 2003. The role of mannose-binding lectin in health and disease. *Mol Immunol,* 40**,** 423-9.

TYCKO, J., MYER, V. E. & HSU, P. D. 2016. Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. *Mol Cell,* 63**,** 355-70.

UHLEN, M., FAGERBERG, L., HALLSTROM, B. M., LINDSKOG, C., OKSVOLD, P., MARDINOGLU, A., SIVERTSSON, A., KAMPF, C., SJOSTEDT, E., ASPLUND, A., OLSSON, I., EDLUND, K., LUNDBERG, E., NAVANI, S., SZIGYARTO, C. A., ODEBERG, J., DJUREINOVIC, D., TAKANEN, J. O., HOBER, S., ALM, T., EDQVIST, P. H., BERLING, H., TEGEL, H., MULDER, J., ROCKBERG, J., NILSSON, P., SCHWENK, J. M., HAMSTEN, M., VON FEILITZEN, K., FORSBERG, M., PERSSON, L., JOHANSSON, F., ZWAHLEN, M., VON HEIJNE, G., NIELSEN, J. & PONTEN, F. 2015. Proteomics. Tissue-based map of the human proteome. *Science,* 347**,** 1260419.

VAN POUCKE, S. 2013. *Determinants in the pandemic H3N2 influenza A virus hemagglutinin that affect receptor binding and transmission in the pig.* info:eu-repo/semantics/publishedVersion, Ghent University. Faculty of Veterinary Medicine.

VARKI, A. 2008. Sialic acids in human health and disease. *Trends Mol Med,* 14**,** 351-60.

VASUDEVAN, D. & HALTIWANGER, R. S. 2014. Novel roles for O-linked glycans in protein folding. *Glycoconj J,* 31**,** 417-26.

VERHAGEN, J. H., LEXMOND, P., VUONG, O., SCHUTTEN, M., GULDEMEESTER, J., OSTERHAUS, A., ELBERS, A. R. W., SLATERUS, R., HORNMAN, M., KOCH, G. & FOUCHIER, R. A. M. 2017. Discordant detection of avian influenza virus subtypes in time and space between poultry and wild birds; Towards improvement of surveillance programs. *PLoS One.*

VISEL, A., BLOW, M. J., LI, Z., ZHANG, T., AKIYAMA, J. A., HOLT, A., PLAJZER-FRICK, I., SHOUKRY, M., WRIGHT, C., CHEN, F., AFZAL, V., REN, B., RUBIN, E. M. & PENNACCHIO, L. A. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature,* 457**,** 854.

VOUILLOT, L., THELIE, A. & POLLET, N. 2015. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3 (Bethesda),* 5**,** 407-15.

WALL, R. J., POWELL, A. M., PAAPE, M. J., KERR, D. E., BANNERMAN, D. D., PURSEL, V. G., WELLS, K. D., TALBOT, N. & HAWK, H. W. 2005. Genetically enhanced cows resist intramammary Staphylococcus aureus infection. *Nat Biotechnol,* 23**,** 445-51.

WANG X , E. A. 1993. Chromosome mapping and organization of the human beta-galactoside alpha 2,6-sialyltransferase gene. Differential and cell-type specific usage of up... - PubMed - NCBI.

WATSON, D. C., WAKARCHUK, W. W., GERVAIS, C., DUROCHER, Y., ROBOTHAM, A., FERNANDES, S. M., SCHNAAR, R. L., YOUNG, N. M. & GILBERT, M. 2015. Preparation of legionaminic acid analogs of sialo-glycoconjugates by means of mammalian sialyltransferases. *Glycoconj J,* 32**,** 729-34.

WEINSTEIN, J., LEE, E. U., MCENTEE, K., LAI, P. H. & PAULSON, J. C. 1987. Primary structure of beta-galactoside alpha 2,6-sialyltransferase. Conversion of membrane-

bound enzyme to soluble forms by cleavage of the NH2-terminal signal anchor. *J Biol Chem,* 262**,** 17735-43.
WENTE, S. R. & ROUT, M. P. 2010. The Nuclear Pore Complex and Nuclear Transport. *Cold Spring Harb Perspect Biol.*
WHITWORTH, K. M., ROWLAND, R. R. R., PETROVAN, V., SHEAHAN, M., CINO-OZUNA, A. G., FANG, Y., HESSE, R., MILEHAM, A., SAMUEL, M. S., WELLS, K. D. & PRATHER, R. S. 2019. Resistance to coronavirus infection in amino peptidase N-deficient pigs. *Transgenic Res,* 28**,** 21-32.
WHYTE, J. J. & PRATHER, R. S. 2011. Genetic modifications of pigs for medicine and agriculture. *Mol Reprod Dev,* 78**,** 879-91.
WICHERT, B., DEPARTMENT OF GYNECOLOGY, U. M. C. H.-E., HAMBURG, GERMANY, MILDE-LANGOSCH, K., DEPARTMENT OF GYNECOLOGY, U. M. C. H.-E., HAMBURG, GERMANY, GALATENKO, V., DEPARMENT OF MATHEMATICAL ANALYSIS, F. O. M. A. M., LOMONOSOV MOSCOW STATE UNIVERSITY, LENINSKIE GORY 1, MOSCOW, RUSSIA, TAUBER BIOINFORMATICS RESEARCH CENTER, U. O. H., 199 ABA KHOUSHY AVE., MOUNT CARMEL, HAIFA, ISRAEL, SCHMALFELDT, B., DEPARTMENT OF GYNECOLOGY, U. M. C. H.-E., HAMBURG, GERMANY, OLIVEIRA-FERRER, L. & DEPARTMENT OF GYNECOLOGY, U. M. C. H.-E., HAMBURG, GERMANY 2018. Prognostic role of the sialyltransferase ST6GAL1 in ovarian cancer. *Glycobiology,* 28**,** 898-903.
WILKES, T. M., DEVONSHIRE, A. S., ELLISON, S. L. & FOY, C. A. 2010. Evaluation of a novel approach for the measurement of RNA quality. *BMC Res Notes.*
WILKS, S., DE GRAAF, M., SMITH, D. J. & BURKE, D. F. 2012. A review of Influenza haemagglutinin receptor binding as it relates to pandemic properties. *Vaccine,* 30**,** 4369-76.
WILLADSEN, S. M. 1986. Nuclear transplantation in sheep embryos. *Nature,* 320**,** 63-5.
WOOD, A. J., LO, T. W., ZEITLER, B., PICKLE, C. S., RALSTON, E. J., LEE, A. H., AMORA, R., MILLER, J. C., LEUNG, E., MENG, X., ZHANG, L., REBAR, E. J., GREGORY, P. D., URNOV, F. D. & MEYER, B. J. 2011. Targeted genome editing across species using ZFNs and TALENs. *Science,* 333**,** 307.
WOROBEY, M., HAN, G.-Z. & RAMBAUT, A. 2014. Genesis and pathogenesis of the 1918 pandemic H1N1 influenza A virus.
WRIGHT, G., CARVER, A., COTTOM, D., REEVES, D., SCOTT, A., SIMONS, P., WILMUT, I., GARNER, I. & COLMAN, A. 1991. High level expression of active human alpha-1-antitrypsin in the milk of transgenic sheep. *Biotechnology (N Y),* 9**,** 830-4.
WU, D., HUANG, W., WANG, Y., GUAN, W., LI, R., YANG, Z. & ZHONG, N. 2014. Gene silencing of beta-galactosamide alpha-2,6-sialyltransferase 1 inhibits human influenza virus infection of airway epithelial cells. *BMC Microbiol,* 14**,** 78.
WU, W. W., SUN, Y. H. & PANTE, N. 2007. Nuclear import of influenza A viral ribonucleoprotein complexes is mediated by two nuclear localization sequences on viral nucleoprotein. *Virol J,* 4**,** 49.
XIANG, G., REN, J., HAI, T., FU, R., YU, D., WANG, J., LI, W., WANG, H. & ZHOU, Q. 2018. Editing porcine IGF2 regulatory element improved meat production in Chinese Bama pigs. *Cell Mol Life Sci,* 75**,** 4619-4628.
XU, R., MCBRIDE, R., PAULSON, J. C., BASLER, C. F. & WILSON, I. A. 2010. Structure, receptor binding, and antigenicity of influenza virus hemagglutinins from the 1957 H2N2 pandemic. *J Virol,* 84**,** 1715-21.
YAN, Q., YANG, H., YANG, D., ZHAO, B., OUYANG, Z., LIU, Z., FAN, N., OUYANG, H., GU, W. & LAI, L. 2014a. Production of transgenic pigs over-expressing the antiviral gene Mx1. *Cell Regen (Lond).*
YAN, Q., ZHANG, Q., YANG, H., ZOU, Q., TANG, C., FAN, N. & LAI, L. 2014b. Generation of multi-gene knockout rabbits using the Cas9/gRNA system. *Cell Regen (Lond),* 3**,** 12.
YAN, S., TU, Z., LIU, Z., FAN, N., YANG, H., YANG, S., YANG, W., ZHAO, Y., OUYANG, Z., LAI, C., LI, L., LIU, Q., SHI, H., XU, G., ZHAO, H., WEI, H., PEI, Z., LI, S., LAI, L. & LI, X. J. 2018. A Huntingtin Knockin Pig Model Recapitulates Features of Selective Neurodegeneration in Huntington's Disease. *Cell,* 173**,** 989-1002.e13.

YANO, K., MOROTOMI-YANO, K., WANG, S. Y., UEMATSU, N., LEE, K. J., ASAITHAMBY, A., WETERINGS, E. & CHEN, D. J. 2008. Ku recruits XLF to DNA double-strand breaks. *EMBO Rep,* 9**,** 91-6.

YP HU, M. D., JTY LAU 1997. Murine hepatic β-galactoside α2,6-sialyltransferase gene expression involves usage of a novel upstream exon region | SpringerLink.

YU, X., LIANG, X., XIE, H., KUMAR, S., RAVINDER, N., POTTER, J., DE MOLLERAT DU JEU, X. & CHESNUT, J. D. 2016. Improved delivery of Cas9 protein/gRNA complexes using lipofectamine CRISPRMAX. *Biotechnol Lett,* 38**,** 919-29.

ZABOIKIN, M., ZABOIKINA, T., FRETER, C. & SRINIVASAKUMAR, N. 2017. Non-Homologous End Joining and Homology Directed DNA Repair Frequency of Double-Stranded Breaks Introduced by Genome Editing Reagents. *PLoS One,* 12**,** e0169931.

ZAIA, J. 2010. Mass Spectrometry and Glycomics. *OMICS.*

ZENG, H., GOLDSMITH, C. S., MAINES, T. R., BELSER, J. A., GUSTIN, K. M., PEKOSZ, A., ZAKI, S. R., KATZ, J. M. & TUMPEY, T. M. 2013. Tropism and Infectivity of Influenza Virus, Including Highly Pathogenic Avian H5N1 Virus, in Ferret Tracheal Differentiated Primary Epithelial Cell Cultures.

ZHANG, J. P., LI, X. L., LI, G. H., CHEN, W., ARAKAKI, C., BOTIMER, G. D., BAYLINK, D., ZHANG, L., WEN, W., FU, Y. W., XU, J., CHUN, N., YUAN, W., CHENG, T. & ZHANG, X. B. 2017a. Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biol.*

ZHANG, S., LU, J., XU, Z., ZOU, X., SUN, X., XU, Y., SHAN, A., LU, J., YAN, X., CUI, Y., YAN, W., DU, Y., GU, J., ZHENG, M., FENG, B. & ZHANG, Y. 2017b. Differential expression of ST6GAL1 in the tumor progression of colorectal cancer. *Biochem Biophys Res Commun,* 486**,** 1090-1096.

ZHENG, Q., LIN, J., HUANG, J., ZHANG, H., ZHANG, R., ZHANG, X., CAO, C., HAMBLY, C., QIN, G., YAO, J., SONG, R., JIA, Q., WANG, X., LI, Y., ZHANG, N., PIAO, Z., YE, R., SPEAKMAN, J. R., WANG, H., ZHOU, Q., WANG, Y., JIN, W. & ZHAO, J. 2017. Reconstitution of UCP1 using CRISPR/Cas9 in the white adipose tissue of pigs decreases fat deposition and improves thermogenic capacity. *Proc Natl Acad Sci U S A,* 114**,** E9474-e9482.

ZURIS, J. A., THOMPSON, D. B., SHU, Y., GUILINGER, J. P., BESSEN, J. L., HU, J. H., MAEDER, M. L., JOUNG, J. K., CHEN, Z.-Y. & LIU, D. R. 2014. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing. *Nature Biotechnology,* 33**,** 73.